THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 57, PART 5 1 September 1945 No. 323

# THE FRICTIONLESS STATE OF AGGREGATION

BY K. MENDELSSOHN,
Clarendon Laboratory, Oxford

*MS. received* 26 *March* 1945

*ABSTRACT.* The analogy of superconductivity and liquid helium II has been analysed on a purely empirical basis. Frictionless transport is considered to be the fundamental phenomenon. It has been concluded that the " superconductive " electrons and the " superfluid " helium atoms form an aggregate in momentum space of zero thermal energy (*z-state*). The ***z-particles*** occupy a set of lowest quantum states which is separated from the thermal states by an energy gap; they have zero entropy, but an appreciable zero-point energy. The observed specific-heat anomalies are due to a lifting of z-particles over this gap. The transport phenomena are caused by a diffusion of z-particles under their zero-point momentum. This zero-point diffusion does not result in dissipation of momentum. It explains why frictionless transport is independent of the strength of external forces and explains also the critical transport rates (threshold value and film transfer). Bose-Einstein condensation fails to account for these characteristic features because it does not lead to a zero-point energy of the condensed phase. A number of experiments to test the hypothesis are suggested.

## § 1. INTRODUCTION

SOME years ago J. G. Daunt and the author (1942) drew attention to the strong similarity between the phenomena of superconductivity and liquid helium II, which becomes apparent when we consider quite generally the properties of particle transport in these substances. In both cases we encounter *frictionless* transport which obeys similar rules in spite of the fact that the particles are atoms in one case and electrons in the other. This, together with an equally striking similarity in the entropy functions (Daunt and Mendelssohn, 1945) has led us to advance the hypothesis that both phenomena are caused by much the same mechanism. We have described this mechanism as the appearance, at finite temperature, of particles which are energetically at absolute zero and do not exchange momentum with the remainder of the substance. In view of the peculiar way in which the thermally unexcited particles are separated from those in the higher quantum states—they do not form a separate phase—we have described the assembly of these particles as an aggregation.

It is the aim of this paper to discuss in more general terms * the hypothesis of the basic identity of superconductivity and liquid helium, and the properties of the aggregation of particles at the lowest quantum levels. It must be emphasized that these considerations are not based on any particular theoretical conception but that an attempt is made to deduce the properties of the frictionless state empirically on the basis of the existing experimental evidence.

---

* A detailed discussion of the energy state of the electrons in superconductors has recently been given elsewhere (Daunt and Mendelssohn, 1945) and a detailed account of the corresponding phenomena in liquid helium will be published later.

## §2. FRICTIONLESS TRANSPORT

Let us first consider the position as regards the experimental facts.* In the case of liquid helium the last years before the war had brought to light a considerable number of seemingly unconnected phenomena, the most striking of which are: extremely low viscosity, very high heat conduction and, in some way related to the latter, a thermo-mechanical effect (sometimes called the "fountain effect") which constitutes a flow of helium in the direction of the higher temperature. In addition there are the transport phenomena in a film of thickness about $5 \times 10^{-6}$ cm. which covers all solid surfaces in contact with liquid helium II. The somewhat unsatisfactory state of the experimental material is due to the fact that internal friction and heat transport in liquid helium II are evidently very complicated processes. The efforts to determine such properties as viscosity and heat conduction have consequently led to complex, and in many cases contradictory, results. On the other hand, the film transfer which at first sight appeared to complicate matters has proved to be of great help since the phenomena in films seem to present a much simplified picture of the transport effects in the bulk liquid and have the added advantage of yielding well reproducible results.

With regard to superconductivity, the position is more satisfactory on account of the greater wealth of confirmed experimental results. Since 1933, when it was discovered that zero resistance is coupled with zero magnetic induction, there has been a tendency to describe superconductivity as a purely magnetic phenomenon, some authors going so far as to consider the disappearance of resistance to be due to complete diamagnetic polarization. Such an explanation is, however, hardly tenable in view of the recent determination of the Landé factor by Kikoin and Goobar (1940), which shows that the super-currents are caused by movement of free electrons and are not merely simulated by electron spin. We stress this point since it will appear from our discussion that superconductivity is primarily an electric and not a magnetic phenomenon. The connection between electric and magnetic effects in superconductors has been expressed in rigorous form by the work of F. and H. London (1935). Their fundamental equation which connects the magnetic field with the current density does not, however, imply, as has sometimes been assumed (Burton *et al.*, 1940, p. 320), that the magnetic effects are the primary ones. The London theory is a phenomenological one which does not differentiate between cause and effect.

We have already given a short summary of the corresponding phenomena in superconductivity and liquid helium (Daunt and Mendelssohn, 1942), and it is not intended to discuss the analogy here in greater detail. Our hypothesis implies that the fundamental phenomenon in both cases is that of *frictionless transport*, and we shall now consider the experimental justification for this statement. In superconductors the evidence is provided by the classical current-potential measurement (figure 1 *a*) in which part of an electric circuit is superconductive. An ammeter in the circuit indicates a flow of electrons through the superconductor while a voltmeter shows zero potential between any two places of the superconductive part of the circuit.* Thus we have a transport of particles—electrons—through the superconductive metal in which no energy is dissipated. The corresponding effect in liquid helium II is illustrated by the transfer of helium atoms in a film over the wall of a beaker (Daunt and Mendelssohn, 1939 a) (figure 1 *b*). If the level inside the beaker is higher than outside, helium atoms are transferred out of the beaker until the difference in level has disappeared. The most remarkable fact about this transfer is that in contradistinction to the action of a syphon the rate of transfer is independent of the length of the path and of the difference in the height of the levels. This goes so far that when the level difference is suddenly changed, no change at all is observed in the rate of transfer. Here again we have a transport of particles—in this case helium atoms—without friction, for friction would make the rate of transfer dependent both on the difference in potential and on the length of the path which the atoms have to traverse. The same phenomenon has been observed in flow of the bulk liquid through capillaries (Allen and Misener, 1939 ; Johns, Wilhelm and Smith, 1939) although not so clearly defined because it is partly masked by viscous flow of the liquid as a whole. However, in very fine capillaries ($\sim 10^{-5}$ cm.) the result is practically identical with the observations on films, and for wider capillaries it can be expressed in the form

---

* Summaries of the phenomena in liquid helium have been given by Darrow (1940) and Jones (1940). A description of superconductivity has been given by Shoenberg (1938) and a summary of recent work by Jackson (1940).

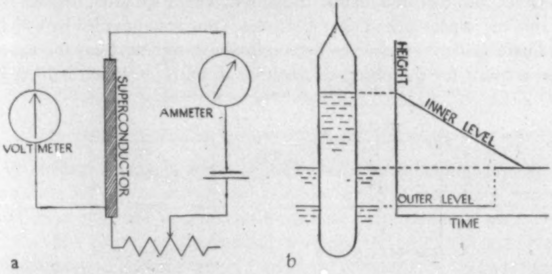


Figure 1. Experiments showing frictionless transport, (*a*) in a superconductor, and (*b*) in a film of liquid helium II. A sudden change in the level difference does *not* affect the rate of transfer.

$$r = R + n\Delta P, \qquad \ldots\ldots (1)$$

where $r$ is the total rate of transport (volume/time), $R$ the rate due to frictionless flow and $n\Delta P$ the amount transported by viscous flow under the pressure difference $\Delta P$. The value of $R$ and its dependence on temperature in the finest capillaries have been found to be of the same order as in films. This means that, as in superconductors, we are faced with the fact of frictionless transport through the rest of the substance. There is some evidence that both in superconductors and in liquid helium the frictionless transport shows a preference for the geometric surface of the substance—a phenomenon which will be discussed later, and which does not affect our general argument.

---

* The most accurate determinations (using persistent currents) show that the resistivity of a superconductor is less than $10^{-20}$ ohm-cm.

The experiments thus show that there are "superconductive" electrons and "superfluid" helium atoms which move through their respective "container substances" without friction. When using such terminology we must, however, keep in mind that we cannot distinguish between individual electrons or individual helium atoms, and that these terms can only be used in allotting a certain number of particles to a given set of energy states. In the present case the particular set of states is distinguished by the fact that the electrons or helium atoms in this set—we will call them *z-particles* for short—do not dissipate momentum.

It must be pointed out straight away that the z-particles do not form a separate phase in either case. Apart from the lack of any direct experimental evidence to this effect, the thermodynamic functions, which we shall discuss later, show clearly that the appearance of the z-particles is not accompanied by a phase change in coordinate space. It has even been impossible to detect any increase in spatial order to account for the observed decrease in entropy which is linked with the appearance of the z-particles. There is, however, another way in which the assembly of z-particles is distinguished from any other assembly of spatially disordered, freely mobile particles belonging to a particular set of states. This is shown by figure 2, which illustrates diagrammatically the relation between the z-particles, the "container substance" and an external force (magnetic field, gravitation) acting on* the z-particles. There exists no interaction between the z-particles and the rest of the substance, but a strong interaction between them and the external force, which means that the z-particles react to it as a *separate aggregate*. More than this, the z-particles assume under the influence of such an external force the aspect of a spatially ordered aggregate whose degree of order can become manifest in macroscopic dimensions (e.g. the magnetic field of a persistent current). It seems therefore convenient to describe the z-particles as forming a particular state of aggregation—since we wish to avoid the term *phase*—which we will call the frictionless state of aggregation (z-state).

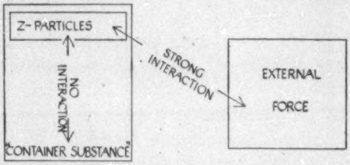


Figure 2: The peculiar position of the "aggregate" of z-particles; they exhibit no interaction with the rest of the substance, but strong interaction with an external force.

It must be stressed that this term does not of course refer to the superconductive metal or liquid helium as a whole but only to the assembly of z-particles, i.e. the "superconductive" electrons and the "superfluid" helium atoms. As the temperature of the substance is raised from absolute zero, the z-state weakens and finally disappears. We will show in § 3 that this phenomenon can be described as a gradual decrease in the number of z-particles with increasing temperature. The frictionless state shrinks until, at the transition point or the λ-point respectively, the last z-particle disappears. In superconductors an indication of the change in population of the z-state is given by the threshold curve which measures the number of electrons that can be transported without friction in a given time over a superconductive surface of given dimensions at different temperatures. We have determined a similar quantity for the frictionless transport of helium atoms in films (Daunt and Mendelssohn, 1939a). It is significant that this "rate of transfer"—the number of helium atoms transported without friction in unit time over a helium surface of given dimensions at different temperatures—was found to depend on temperature only. Figure 3 gives a typical threshold curve and the rate of transfer in comparison. These functions, which merely give the amount of z-particles collected in unit time, contain of course an unknown average velocity. If this velocity is independent of the temperature, as appears likely (see § 5), the curves can be regarded as indicating directly the relative number of z-particles as a function of the temperature.

---

* As will be seen in § 5 the conception of an "accelerating" force has to be qualified, since it seems possible that all frictionless transport is due to zero-point momentum, the acting force providing space order rather than acceleration.

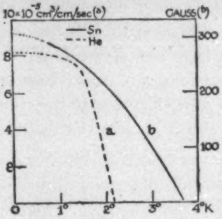


Figure 3. Critical transport rates (*a*) in liquid helium, and (*b*) in a superconductor (tin).

## § 3. THERMAL ENERGY OF THE FRICTIONLESS STATE

The gradual disappearance of the frictionless state of aggregation is accompanied both in superconductors and in liquid helium II by an excess specific heat (figure 4). In superconductors it has been possible to separate the specific heat which is due to the electrons from that caused by the lattice vibrations (Daunt, Horseman and Mendelssohn, 1939). It has sometimes been assumed

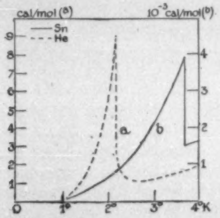


Figure 4. Specific heat (*a*) of liquid helium, and (*b*) of the free electrons in tin.

that this excess specific heat, which is caused somehow by those particles participating in the anomalous state, must imply heat exchange between a system of these particles and the rest of the substance. The point has been discussed in the case of superconductors (Burton *et al.*, 1940, p. 298). The mechanism sketched in figure 2 on the other hand makes it very difficult to account for such an interaction. In fact, our definition of the z-particles—and this definition was derived from experimental results—would not allow direct thermal interaction with the " container substance " unless we provide for it by additional assumptions. Any such assumption, allowing energy exchange in one case and prohibiting it in another, would moreover be highly artificial.

This obvious discrepancy requires an explanation and it raises the question of the energy state of the z-particles, i.e. of the heat content of the z-state. The specific heats given in figure 4 refer of course to the contributions by the z-particles *and* by the normal particles, and it is clear that information on the thermal state of the z-particles by themselves must be obtained by experiments of a different character.

Such experiments have been carried out both for a superconductor and for liquid helium II. In the superconductor this is the determination of the Thomson heat (Daunt and Mendelssohn, 1938); the experiment and the result have been discussed in detail elsewhere (Daunt and Mendelssohn, 1945). The fact that the Thomson heat of a persistent current is zero means that the superconductive electrons do *not* exchange thermal energy with the rest of the substance. This leads to a peculiar consequence when we consider what must happen in a ring of superconductive metal with a persistent current running in it, which is warmed up from absolute zero. Since the electrons in the current are not in heat exchange with the lattice, they must remain energetically at absolute zero although the substance as a whole warms up.

An idea as to the thermal state of the superfluid helium atoms can be obtained from any experiment which causes their separation in space from the liquid as a whole. We have seen in the last section that, while there exists no space order generally, an external force has such an ordering influence (see figure 2). In determining the Thomson heat in a superconductor we have therefore made use of the influence of the magnetic field, and in the case of helium we can use gravitation as the acting force. In the experiment shown in figure 1 *b* superfluid helium atoms were separated by surface flow. If the heat content of the superfluid helium atoms, like that of the superconductive electrons, is smaller than that

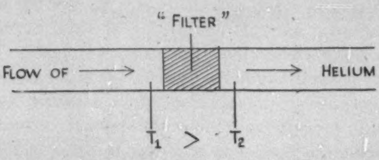


Figure 5. The mechano-caloric effect in which a separation of z-particles occurs. A flow of liquid helium II through a powder " filter " is accompanied with a heating of the liquid at the entrance and a cooling at the exit.

of the normal particles, the liquid drained off must be colder than that inside the beaker. We have carried out an experiment of this type (Daunt and Mendelssohn, 1939 b), using for practical reasons the flow of liquid helium II through tightly packed fine powder instead of flow over an exposed surface. A diagrammatical sketch of the method is given in figure 5. As was expected, a heating was observed at the entrance of the liquid into the "filter" and a cooling at the exit. H. London (1939) has since given a general thermodynamical interpretation of this mechano-caloric effect. Under the assumption of complete reversibility he obtains the following relation between the pressure ($p$) under which the liquid is forced through the power plug and the entropy difference ($\Delta S$) between the helium flowing without friction and the bulk liquid :

$$\frac{dp}{dT} = \rho \Delta S, \qquad \ldots\ldots (2)$$

where $\rho$ is the density of the liquid. Our experiments (which had to be suspended in 1939) were of a preliminary character only, but the ($dp/dT$)-values obtained [*] yielded a large entropy change indicating that the superfluid helium atoms seemed to have lost *all* thermal energy.[†] Fortunately the experiment has since been repeated by Kapitza (1941), using a narrow gap instead of the powder plug. He could not only confirm the existence of the mechano-calorific effect and demonstrate with greater accuracy the complete loss of thermal energy suffered by the superfluid atoms, but by measuring independently the heat change connected with the effect, he has shown that the process is reversible.

It has to be admitted that we cannot yet establish the complete loss of thermal energy with the same accuracy for the superfluid helium atoms as in the case of the superconductive electrons.[‡] Particularly the specific heat at low temperatures (Pickard and Simon, 1945) has been interpreted as corresponding to a "lattice entropy" (*cf.* H. London, 1939) while frictionless transport has been assumed at the same time. We will return to this question, which is not at all clear, in § 4. For the present we shall neglect this hypothetical residual entropy which, as the experiments show, does not interfere with frictionless transport.

The experimental result that in a "container substance" which has a finite temperature, the z-particles remain energetically at absolute zero but have momentum, leads to the simple but somewhat surprising conclusion that they must move without friction. For friction is the dissipation of momentum and any particle engaged in a process involving friction must therefore suffer thermal excitation. This means that the observation that z-particles remain thermally unexcited while being capable of translatory motion is in itself sufficient to explain the phenomenon of frictionless transport. It seems permissible to speak of the entropy of the z-state since this state, as will be seen, forms a separate phase in momentum space, and it is clear that this entropy could not retain the value zero—as we know it does—when the state is agitated, unless this agitation does not involve friction.

---

[*] The work will be published in detail later.

[†] Allen and Reekie (1939 *a*) have obtained ($dp/dT$)-values from the thermo-mechanical (fountain-) effect which is the thermodynamic reverse of the mechano-caloric effect. These values are of the same order as those from the mechano-caloric effect, but again not accurate enough to determine the exact entropy loss.

[‡] Our conception of the properties of the z-state suggests that when a rotating mass of liquid helium is cooled below the $\lambda$-point, a frictionless current of z-particles will be set up which will persist when the momentum of the thermally excited particles has been dissipated. The decay of this persistent helium current will provide an equally sensitive measure for the frictional resistance of the superfluid atoms as its electric counterpart does for the resistances of the flow of superconductive electrons.

The realization that the entropy of the frictionless state is zero at once removes the discrepancy with which we were faced when considering the excess specific heat of the z-particles. It also answers the question as to the significance of the entropy change accompanying the weakening of the frictionless state with increasing temperature. There exists no direct thermal interaction between the z-particles and the rest of the substance—the specific heat of the z-state is zero. The observed specific heat is clearly due to an excitation process in which z-particles are lifted into thermally excited states. When this occurs, they cease being z-particles and cease to be available for frictionless transport.

The anomaly in the specific heat connected with this change (figure 4) exhibits the familiar pattern associated with "cooperative" phenomena like ferromagnetism, rochelle-electricity and order-disorder transformations. The latter in particular have been made the subject of detailed statistical treatment, the methods of which have been applied to other phenomena, as for instance molecular rotation in solids. However, no general statistical treatment has yet been developed which can be rigorously applied to all these changes without regard to the particular physical process involved. Simon (1930) who first investigated these "quantum jumps" in solids has generalized them as excitations of internal degrees of freedom. The large amount of experimental data gathered in the last twenty years exhibits all stages of specific-heat anomalies ranging from the uninfluenced Schottky curve over λ-points to transitions involving a latent heat. This fact is important for our particular problem, since it shows that no clear-cut distinction is possible with regard to the mechanisms of the various excitation processes, and that to describe an unknown physical process on the strength of the observed specific-heat function as an order-disorder transformation

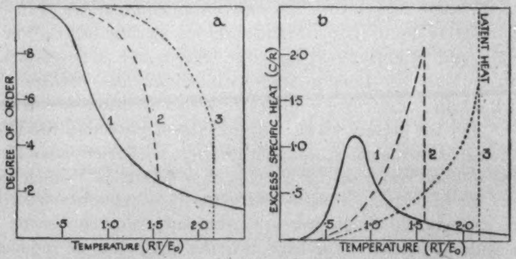


Figure 6. (*a*) Degree of order, and (*b*) excess specific heat of different alloys (theoretical). (1) Type AB with short range order, (2) type AB with long range order, and (3) type $AB_3$ with long range order. $E_0$ is the total energy of the transformation required to change the alloy from complete order to complete disorder. (Cf. Fowler (1936); Nix and Shockley (1938).)

or a rotational change has little meaning. It usually means only that the particular statistical treatment can be applied, but it discloses nothing with regard to the *nature* of the process. In the case of the order-disorder transformations it has been possible to connect the theoretically determined states of order with specific-heat functions, which approach the observed ones, for a non-cooperative (short-range order, AB) type,* a cooperative (long-range order, AB) type, and a highly cooperative ($AB_3$) type. Figure 6 illustrates the point that the type of the change is not indicative of its nature. The degree of order—the inverse of the entropy—is a function, and in certain cases a direct measure, of the population of the lower state.

Considering the apparent lack of space order and the high symmetry of both the electron and the helium atom, we are forced to assume that the degrees of freedom lost by the assembly of z-particles are those of translatory energy. There is nothing unusual in the fact that at finite temperatures some particles in a substance are energetically at absolute zero. This is actually true for a proportion of the atoms in any solid at low enough temperatures. However, in the case of the z-state we deal with freely mobile particles. The nature of the phenomenon of frictionless transport is therefore the manifestation of an effect predicted by Nernst (1926) when he extended the third law to systems with freely mobile particles, i.e. gas degeneracy.

## §4. THE ENERGY SPECTRUM

While thus frictionless transport is connected with gas degeneracy, this does not mean that it is a necessary consequence of it. In fact, the free electrons in the Sommerfeld model are in a highly degenerate state without exhibiting frictionless transport. The existence of z-particles, capable of translatory motion but retaining zero entropy, requires therefore not only degeneracy but degeneracy of a particular type. The particular qualification necessary for frictionless transport becomes apparent when we consider the term system of the Sommerfeld model (figure 7 *a*). A continuum of energy states is filled at absolute zero up to the level (L) which divides a completely filled set of states (A) from a completely unoccupied one (B). At finite temperatures an accelerated electron can and will dissipate energy, since it can be lifted to *any* place above (L). If we wish to allow for frictionless transport, we must prohibit states bordering (L), so that no dissipation of momentum can take place ; i.e. we must introduce a *gap* in the term system (figure 7 *b*). As we have pointed out elsewhere (1945), in superconductors this gap forms the limit of the Fermi distribution at absolute zero. It follows from the considerations discussed in the preceding sections that such a term system in which the continua (A) and (B) are separated by a gap in the energy spectrum must be characteristic not only of a superconductor but of any other z-state. (The only other case known is, of course, liquid helium II.)

The existence of the gap signifies a separation in momentum space between the particles in (A) which cannot dissipate momentum and those in (B) which are thermally excited and acceleration of which results in friction since they can

* This function, a Schottky curve, has been observed by Simon (1926) in a number of pure metals ; the nature of the change is unknown. Other examples of this type of anomaly are the paramagnetic salts and solid ortho-hydrogen.

be lifted to any higher state in the continuum (B). We can therefore speak of a "condensation" in momentum space, the z-particles forming the "condensed" phase. This phase forms the frictionless state of aggregation. Neither in superconductors nor in liquid helium is this a condensation to zero energy in the model sketched above. This means that neither case corresponds to a Bose-Einstein condensation since both have an appreciable zero-point energy. As mentioned in § 3, it is by no means clear what happens to the "condensed" helium atoms in the liquid. The assumption that the rise in the total entropy of liquid helium below 1° K., which Pickard and Simon found to be proportional to $T^3$, is due to "lattice vibrations" can be brought in accordance, though with

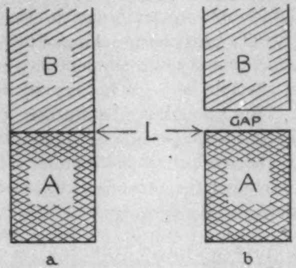


Figure 7. Term systems (*a*) of the electrons in a normal metal, and (*b*) as suggested for the electrons in a super-conductor or the atoms in liquid helium. $A$ is the zero-point energy.

some difficulty only, with the phenomenon of frictionless transport, since it may be argued that the vibrations excited at these temperatures refer to large aggregates of atoms. On the other hand, the available data on the mechano-caloric effect seem to indicate that the *total* entropy disappears,* which would mean that we cannot separate the entropy of liquid helium into an anomalous and a normal part and that even at fairly low temperatures the overwhelming part of the specific heat is made up of the excitation of z-particles. More experimental data are clearly needed to settle this point. Apart from accurate determinations of the mechano-caloric effect, the necessary information can probably be obtained by measuring the coefficient of expansion below 1° K. The observed decrease in density with decreasing temperature is an anomalous effect due to "condensation" of z-particles and if at the lowest temperatures "lattice vibrations" are predominantly responsible for the entropy change, the coefficient of expansion should again change its sign.

The suggested term system removes the difficulty, which has sometimes been emphasized in the case of superconductors, that the z-particles appear to be not in *thermal* equilibrium with the rest of the substance. Such a state would of course not be stable. The reason why they remain energetically at absolute zero even if the container substance has finite temperatures, is simply due to a lack of heat quanta large enough to lift z-particles over the gap. As the temperature of the substance is raised, the proportion of large quanta increases until the gap becomes insignificant. The shape of the specific heat curves (figure 4) indicates that the transition is not uninfluenced by the number of particles in the higher state, since it does not follow a Schottky function (figure 6*b*, 1), but indicates a cooperative phenomenon like ferromagnetism or the influence of long-range order on an order-disorder transformation. The rough term system given in figure 7*b* does not, of course, allow conclusions to be drawn as to the shape of the specific-heat curve, because it is completely empirical and no assumptions are made about the density of states above or below the gap. It must be emphasized that, since our considerations are based entirely on experimental observations, no theoretical explanation for the existence of such gaps in the energy spectrum can be given.

---

* At 1°·5 K., for instance, the observed difference between the entropy of the z-particles and the entropy of the whole liquid ($\Delta S$ in formula (2)) is $5 \times 10^{-2}$ cal./g., while the total entropy has been calculated (Simon, unpublished) as $5{\cdot}04 \times 10^{-2}$ cal./g. The extrapolated "lattice entropy" ($\theta = 15$, Pickard and Simon, unpublished) of $3{\cdot}8 \times 10^{-3}$ cal./g. is too big to be accommodated in the difference between the first two figures.

## §5. ZERO-POINT DIFFUSION

So far our account has given no explanation of two outstanding characteristics of the z-state, the breakdown of frictionless transport when a critical rate of transport is exceeded and the high heat-conduction of liquid helium II. Both phenomena are, as we shall see, intimately connected. When investigating the high heat-conduction of the helium film (Daunt and Mendelssohn, 1939 a), we were able to show that this also was due to a transport phenomenon, the flow of super-fluid helium atoms towards the place where heat was supplied. We then suggested that the heat conduction in the bulk liquid would probably be due to a similar though rather more complicated convection process. That this is actually the case has since been demonstrated by Kapitza (1940) in a number of experiments showing the existence of a counter current of liquid helium in a capillary under a temperature gradient. For an investigation of the mechanism of the process let us revert to the simple case of the helium film (figure 8).

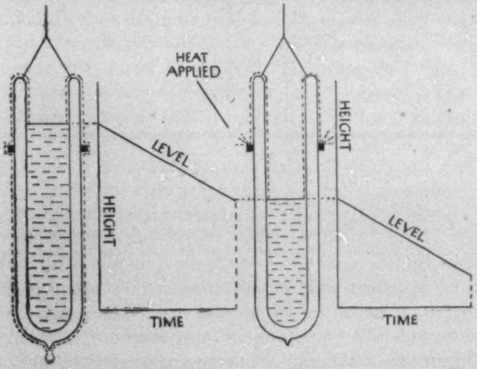


Figure 8. The rate of flow surface of liquid helium II is unaffected by the application of heat to the film. This can be explained by the assumption that the z-particles diffuse under their zero-point momentum.

We observed that the rate at which helium is transferred out of a Dewar vessel is *not* affected when heat is applied to the film outside the vessel. This not only demonstrates that the observed high heat-transport through the film is entirely due to a transport of helium atoms, but also that this process must evidently be *isothermal*, since a heating of the film at one place would have resulted in heat flow into the vessel by conduction. The salient features of the film transfer discussed above (§2) show that only z-particles are engaged in the movement and the heat supplied to the film must be taken up as heat of excitation which lifts the z-particles into the continuum of energy states and evaporates them. Heat is of course gained by the helium inside the vessel owing to loss of z-particles, but the *same* gain of heat occurs whether the film is heated or simply allowed to drop off. This explanation could be proved by an experiment (Daunt and Mendelssohn, 1939 b) in which heat was supplied to the inside of a Dewar vessel where the concentration of z-particles was thus decreased (figure 9). In this case a transfer of z-particles *into* the vessel must take place, resulting, as was actually observed, in a rise of the liquid level inside the vessel. The observed effect clearly corresponds to the "fountain phenomenon" discovered by Allen and Jones (1938) in the bulk liquid. The comparatively simple conditions of our experiment now allow an interpretation of the whole set of effects connected with the heat transport in liquid helium II.

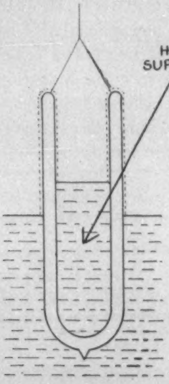


Figure 9. When heat is supplied to the inside of the vessel, the concentration of z-particles is decreased and zero-point diffusion takes place along the film from the bath into the vessel.

The important point in the experiment shown in figure 8 is that there is no "accelerating force" which drives the z-particles to the place at which heat is supplied. It is a striking feature of the experiments that, as was pointed out (Daunt and Mendelssohn, 1939 a), a transfer always takes place to those parts where the film (i.e. z-particles,) is removed. Since the z-particles do not exchange energy with the rest of the substance, the force distributing them evenly over the available surface must be their own zero-point energy. The heat transport in a film of liquid helium II is thus simply due to the *diffusion of z-particles under their zero-point momentum.*

In the bulk liquid, this zero-point diffusion will manifest itself as the tendency of the z-particles to take up a statistical distribution within the space occupied by the "container substance", and a macroscopic movement of z-particles must take place whenever this statistical distribution, i.e. the thermal equilibrium in the substance as a whole, is disturbed. It may be argued that such a "diffusion" process of a set of particles having zero entropy would be in contradiction to the third law. This is, however, not the case since the substance as a whole has finite entropy, and the experiments on helium II (Kürti and Simon, 1938) show that at very low temperatures, where the total entropy becomes small, the thermal changes associated with the flow phenomenon disappear. It is significant that the assemblies exhibiting the z-state—liquid helium and metal electrons—are distinguished by their high zero-point energy. The fundamental importance of the influence of zero-point energy on the phenomena in helium has been emphasized by Simon (1934), who pointed out that its high zero-point energy is responsible for the existence of *liquid* helium at absolute zero.

Our explanation of the heat transport in helium II is therefore the opposite to the one given by Tisza (1938) on the basis of the hypothesis that liquid helium is a Bose-Einstein gas. He attributes *no* momentum to the unexcited particles and explains the flow of heat as due to an osmotic diffusion of the excited particles in the bulk liquid under a temperature gradient. It is difficult to see how such a model in which the flow of unexcited particles only appears as a compensating displacement current can account for the fountain phenomenon. Quite apart from the fact that this model or a similar one proposed by F. London (1939) can never lead to *isothermal* heat transport, it is in direct contradiction to the experiment of figure 8, since there is no bulk liquid for the diffusion of excited particles. Such a diffusion of excited particles as visualized by Tisza may of course take place as a secondary process in the bulk liquid, and, particularly for cases of large heat flow with which the diffusion by zero-point energy cannot cope, become the preponderant feature.* This is probably the explanation for the peculiar negative fountain effect observed by Allan and Reekie (1939b). The inversion point of the fountain observed by these authors will be due to the equilibrium between the two rates of diffusion. The striking high heat-conduction of the bulk liquid for small heat flow and in particular its dependence on the strength of the heat flow cannot however be explained by Tisza's model, while it agrees well with the process of z-particle diffusion outlined above.

The conception of zero-point diffusion of z-particles immediately leads to an explanation of the critical rate of frictionless transport. Taking again the case of the helium film, it is clear that no rise of temperature will take place as long as the z-particles which are excited by the heat supply can be replaced by diffusion. This rate is limited by the number of z-particles per unit volume (depending on temperature) and their zero-point velocity. Once this rate is exceeded, frictionless transport breaks down.

---

* The establishment of a temperature difference in helium II, therefore, depends on two limiting factors, the rate of heat supply and the removal by osmotic and hydrostatic pressure of excited particles. The latter is no problem in our experiment (figure 8), where they are simply evaporated or drop off, but is important in the bulk liquid. The usual type of heat conduction experiments, employing capillaries closed at one or both ends, is thus bound to give results far too complex for interpretation, since it is quite impossible to separate the two limiting factors. A more thorough discussion of the heat-conduction process in helium II will be given by J. G. Daunt and the author later.

It seems reasonable to assume a similar process of zero-point diffusion for the superconductive electrons, since in superconductors, too, we are faced with a critical rate of frictionless transport, i.e. the threshold value. When applying these considerations to the superconductive electrons we arrive at an interesting conclusion. A current set up by zero-point diffusion of electrons must be essentially different from a normal current since the momentum of the flowing electrons is derived solely from collisions * as in the case of helium atoms. This means that the cause of a normal current, the accelerating force of the electric field, does not appear at all in the process. This peculiar phenomenon of a current in zero electric field is of course exactly what has been observed in superconductors (figure 1 *a*). According to our conception there should be another phenomenon, so far unobserved, by which a super-current is distinguished from a normal one. Since it is set up by collisions, the speed of propagation of a (heat) impulse should be different, the maximum speed of a signal depending on the zero-point velocity of the highest level in A (figure 7 *b*). The super-current will also be distinguished by having (in a conductor of uniform cross-section) uniform charge density.

Thus zero-point diffusion seems to explain the curious analogy from which our considerations originated (1942), i.e. that to the absence of potential differences in a superconductor there corresponds the absence of temperature differences in the helium film. The conception of zero-point diffusion also removes at once the difficulty of understanding the independence of the transport rate of z-particles of the *strength* of external forces, e.g. the pressure head in helium. It is clear that our statement (made in §2) concerning the action of an external force on the z-state has to be qualified. Indeed it seems doubtful whether in *any* of the transport phenomena a true particle acceleration occurs. It appears that the action of external forces on z-particles is rather that of creating a certain type of (space-) order than of accelerating particles.

To illustrate this point we consider an element ($dx$) of a long cylinder on the surface of which frictionless transport of $z$-particles can occur. This model corresponds to a superconductive wire or to a solid rod covered with the helium film. If at some place outside ($dx$), z-particles are removed, zero-point diffusion must take place and a macroscopic flow of z-particles in the $x$-direction will occur. This process entails no actual acceleration of z-particles but merely a greater mean free path in the $x$-direction. The flow of z-particles is accompanied in the case of electrons by the appearance of a magnetic (but not an electric) field and in the case of helium atoms by a moment of inertia. The states of flow and rest are thus energetically different but, as the experiments show, of equal (zero) entropy. If we make ($dx$) part of a closed ring, the state of flow, if once started, will persist since there are no means of energy dissipation and we arrive at a metastable state accompanied in superconductors by a persistent magnetic moment and in helium II by a persistent moment of inertia. The former phenomenon is well known, and it should be possible to detect the latter, particularly if these considerations can be extended to a tube filled with the bulk liquid.

---

* When we use here the ordinary kinetic picture of momentum transfer by collision, it is merely done to distinguish the process of zero-point diffusion from that of acceleration by a macroscopic field of force. It has, however, to be remembered that zero-point diffusion is bound to differ from an ordinary diffusion process by the fact that while momentum is exchanged, it is not dissipated, and it is doubtful whether the terms " mean free path " and " average velocity " can be used in their accepted meaning. Retaining such a kinetic interpretation of zero-point energy, it appears that in systems with high zero-point energy there must be a considerable dispersion of velocities. An indication of this seems to have been observed by Ganz (1940) when he found that the first indication of a heat impulse travelling through liquid helium was propagated with a speed of $\sim 10^4$ cm./sec., while the speed of the maximum of the impulse was only $\sim 10$ cm./sec. Incidentally, the secondary maximum observed by him may have been due to Tisza's diffusion of excited particles discussed above.

Another way of describing this phenomenon is to say that the momentum of frictionless transport is not dissipated because it is zero-point energy.

## § 6. SURFACE TRANSPORT

In our 1942 paper, we have drawn attention to the peculiar tendency of the z-particles to move along the geometrical surface of the substance. As a tentative explanation it was suggested that this surface transport may be caused by the fact that the acting force only penetrates the substance to a certain depth and that below this depth no movement of z-particles will take place. It may also be that the disturbance of the structure of the "container substance" produced by the surface facilitates the thermal excitation of z-particles, a process about which nothing is known so far. In this case the depth of penetration should be to some extent independent of the particular type of disturbance and in some way be a measure of the limit of the free path of z-particles.

In superconductors the phenomenon of surface transport is a well-established fact, but in the case of liquid helium the evidence for it is not so strong. In view of the similarity of the transport phenomena in the film with those in the bulk liquid (Daunt and Mendelssohn, 1939 a ; Allen and Misener, 1939) it has been suggested that the film may extend below the liquid level and give rise to a transport of z-particles along the solid walls in the liquid. Subsequent experiments by J. G. Daunt and the author (1939 b), Allen and Reekie (1939 b), and Kapitza (1940) indeed seem to strengthen this view, but it must be pointed out that some of the results can be explained equally well by zero-point diffusion of z-particles in the bulk liquid. A surface flow of superfluid helium atoms which is analogous to the skin effect in superconductors does not however necessitate a higher concentration of z-particles near the walls as is postulated in the model suggested by F. London (1939). A decision as to whether and to what extent the frictionless transport in liquid helium is confined to the walls of the container has to be left to further experiments.

## § 7. BOSE-EINSTEIN CONDENSATION

The conception of the frictionless state of aggregation at which we have arrived in the preceding sections was reached merely by conclusions from the existing experimental evidence and does not contain any theoretical hypothesis. Even so, a fairly consistent model could be obtained, the chief characteristics of which are those of a quantum liquid condensing in momentum space; i.e. a set of freely mobile particles of zero thermal energy but an appreciable zero-point energy which is separated from the continuum of thermal states by an energy gap. In 1938 F. London attempted to explain the $\lambda$ phenomenon of liquid helium in terms of the condensation of a Bose-Einstein gas, and it is interesting to see how far such a model can be used as a theoretical basis for the interpretation of our empirical conception of the z-state.

The chief difficulty is of course that Bose-Einstein statistics is not applicable to electrons in metals. However, as has been pointed out by Lindemann (1932), both Bose-Einstein and Fermi-Dirac statistics refer to idealized cases and it is quite possible that an intermediate treatment may prove equally applicable to electrons and helium atoms (cf. Daunt and Mendelssohn, 1945).

It is evident that the high zero-point energy of both liquid helium and the metal electrons plays a most important part in the behaviour of the z-state whereas the (ideal) Bose-Einstein gas has *no* zero-point energy. This is clearly the reason why Bose-Einstein statistics has failed to provide an explanation for the pressure-independent flow[*] and the isothermal heat transport in helium films. The process of zero-point diffusion which we have introduced in order to explain these phenomena does not exist in the ideal Bose-Einstein gas. Gogate and Rai (1944) have recently compared our rate of transfer of superfluid helium atoms in films (figure 3 *a*) with the velocity[†] of "non-energetic" particles derived from Bose-Einstein statistics and have found a fairly good agreement for the temperature dependency except for the lower temperatures. In their calculation they use the population of the lowest energy state as given by Bose-Einstein statistics to be

$$n' = n[1-(T/T_0)^\sigma], \qquad \ldots\ldots (3)$$

where $n$ is the total number of particles per unit volume, $T_0$ the $\lambda$-point and $\sigma$ has the value 3/2. A much closer agreement can however be obtained, as we pointed out recently (1945), by using the value 5 for $\sigma$, and this is the value substituted by F. London (1939) in his modified theory of Bose-Einstein condensation. It may appear at first as if this remarkable agreement would provide a strong support for the application of Bose-Einstein statistics, but actually the contrary seems to be the case. The value ($\sigma=5$) was chosen by London not on any theoretical grounds, but merely as the value which happens to fit the observed specific-heat curve of liquid helium.* As was shown in § 3, there exists a quite general statistical relation which is entirely independent of the particular type of statistics employed and which connects the population of the lower energy state with the excess specific heat. This relation simply states the fact that the population of the lower state determines the change of the degree of order and with it the excess specific heat in *any* thermal excitation of a degree of freedom. The case has been thoroughly discussed for order-disorder transformations and for molecular rotation (cf. Fowler, 1936). The agreement between the observed rate of transfer and equation (3) has therefore nothing to do with Bose-Einstein statistics, but merely re-states, and incidentally confirms, the experimental fact that the rate of transfer measures the population of the z-state in helium.

---

[*] Tisza has tried to explain this phenomenon by the accompanying thermal effects, while F. London postulates a dependency on the square root of the acting force. Both alternatives have been clearly disproved by the experiments (Daunt and Mendelssohn, 1939 *a*).

[†] In doing this, they tacitly assume the applicability of the quantum-mechanical formula $v \sim \mathbf{h}/md$ for the zero-point velocity of mean free path, where $m$ is the mass of the helium atom and $d$ the mean free path. This is very tempting, particularly since adoption of the film thickness as value for $d$ leads to a velocity value of the same order as the effective velocity observed in the film transport ($\sim 10^2$ cm./sec.). The agreement is, however, deceptive, since the zero-point energy of liquid helium calculated by the same formula gives a value $10^4$ times too small. We have therefore adopted the interpretation of the film thickness as a *limit* (corresponding to the lowest levels of A, figure 7*b*) of free path, given at the end of the first paragraph of § 7 of this paper.

Thus while Bose-Einstein statistics fails to account for the excess specific heat of liquid helium and for the characteristic properties of flow and heat transport, the question arises whether it has any bearing whatsoever on the z-state as it is observed in liquid helium and superconductors. It seems that the position can be summarized as follows : the experiments show that both electrons and helium atoms can pass into a state in which they are capable of frictionless transport. These z-particles are energetically at absolute zero even at finite temperatures—they have zero entropy. The existence of particles in the lowest set of quantum states does not in itself lead to frictionless transport ; it is necessary that these states should be separated from the rest of the energy spectrum by an energy gap. That means that a "condensation" in momentum space must take place if the frictionless state of aggregation is to appear, but this need not be a condensation into a state of no zero-point energy as in Bose-Einstein statistics. According to our considerations (§ 5) zero-point energy is necessary for the appearance of frictionless transport, and in fact no condensation to zero velocity takes place in superconductors and probably not in liquid helium either. The undisputed merit of Bose-Einstein statistics is that, while it cannot be rigorously applied to either superconductors or liquid helium, it is the only statistics so far developed which provides for a condensation phenomenon at all ; and by introducing it F. London was the first to interpret the $\lambda$-phenomenon in terms of gas degeneracy.

The problem of liquid helium has been approached more recently by Landau (1941) who, instead of introducing a specific theoretical model, considers quite generally the quantization of the motion of liquids. He arrives, taking into account the experimental results, at two sets of hydrodynamic equations for the description of liquid helium. In its present state the theory appears still too general to present a rigorous model of liquid helium and it fails to account for the essential features of the transport phenomenon, as for instance isothermal heat transport, pressure-independent flow and the two components of the fountain effect. Its great advantage is that, in contra-distinction to the Bose-Einstein model, it does not require modification of an already existing framework but seems to be capable of application to the case under consideration by mere specialization.

---

* It is interesting that this purely empirical modification, introduced by London to allow for the "unidealty" of his model, constitutes an approach to our suggested term system (figure 7 *b*). It has been shown that the adoption of ($\sigma=5$) instead of ($\sigma=3/2$) seems to produce the postulated gap in the energy spectrum of helium between the set of ground states and the thermally excited states.

## §8. CONCLUSION

The picture of the z-state derived from the experimental facts is far too consistent to permit us to dismiss the similarity between superconductivity and liquid helium II as merely accidental or superficial. Summarizing our conclusions, we arrive at three salient features of the frictionless state of aggregation: (*a*) The z-particles are energetically at absolute zero. (*b*) The zero-point energy is divided by an energy gap from the thermally excited states. (*c*) The transport phenomena are caused by a diffusion of z-particles under their zero-point momentum.

(*a*) is fairly well established by direct experiment. The term system suggested by us to account for (*b*) is a very rough model and we can hardly expect it to yield more than a qualitative explanation, particularly as all evidence for the existence of the gap is indirect. (*c*) is of course closely connected with (*a*) and the fact that frictionless transport is only observed in certain metals is explained by (*b*). (*c*) is strongly suppported by all the available experimental evidence and appears to provide a very satisfactory explanation for the transport phenomena in superconductors and liquid helium. The conception of zero-point diffusion would seem to be the most important conclusion reached in our considerations.

The foremost application of these considerations will be to provide a working hypothesis for experimental research in a field where so far nearly all important progress has been achieved by accidental observations. It is clear that our model can be tested by a great variety of experiments, a number of which has been indicated in this paper. Apart from the application to experimental work it is hoped that the above considerations will also aid the development of a theory of superconductivity and liquid helium. The most serious gap in this respect is the lack of a kinetic interpretation of a system of freely mobile particles with no thermal but finite zero-point energy.

Finally one may ask whether the z-state is realized anywhere in the world except in the laboratory. The high densities in stars and possibly in some of the planets may provide suitable conditions for its appearance in macroscopic dimensions. Such a high concentration of particles also occurs in the nuclei of heavy atoms which have been described by Bohr and Kalckar (1937) as droplets of quantum liquid, and Teller and Wheeler (1938) have drawn attention to a resemblance between the term system of liquid helium and that of the nucleus.

## §9. ACKNOWLEDGMENTS

In the past years the author had occasion to discuss various aspects of the problem with a number of colleagues, and he would like to take this opportunity

of expressing his gratitude to them. He is particularly indebted to Dr. J. G. Daunt, who has helped greatly in the clarification of many points, and to Dr. F. Simon, F.R.S., for providing him with numerical data from unpublished work and for much valuable criticism.

## REFERENCES

ALLEN and JONES, 1938. *Nature, Lond.*, **141,** 243.
ALLEN and MISENER, 1939. *Proc. Roy. Soc.*, A, **127,** 467.
ALLEN and REEKIE, 1939 a. *Proc. Camb. Phil. Soc.* **35,** 114.
ALLEN and REEKIE, 1939 b. *Nature, Lond.*, **144,** 475.
BOHR and KALCKAR, 1937. *Kgl. Danske Videnskab. Selskab,* **14,** no. 10.
BURTON, SMITH and WILHELM, 1940. *Phenomena at the Temperature of Liquid Helium* (New York: Reinold).
DARROW, 1940. *Rev. Mod. Phys.* **12,** 257.
DAUNT, HORSEMAN and MENDELSSOHN, 1939. *Phil. Mag.* **27,** 754.
DAUNT and MENDELSSOHN, 1938. *Nature, Lond.*, **141,** 116; 1939 a. *Proc. Roy. Soc.* A, **170,** 423; 1939 b. *Nature, Lond.*, **143,** 719; 1942. *Nature, Lond.*, **150,** 604; 1945. *Proc. Roy. Soc.* A (in the press).
FOWLER, 1936. *Statistical Mechanics*, 2nd ed. (Cambridge).
GANZ, 1940. *Proc. Camb. Phil. Soc.* **36,** 127.
GOGATE and RAI, 1944. *Nature, Lond.*, **153,** 342.
JACKSON, 1940. *Rep. Progr. Phys.* **6,** 335.
JOHNS, WILHELM and SMITH, 1939. *Canad. J. Res.* **17,** 149.
JONES, 1940. *Rep. Progr. Phys.* **6,** 280.
KAPITZA, 1940. *J. Phys. U.S.S.R.* **4,** 181; 1941. *Phys. Rev.* **60,** 354.
KIKOIN and GOOBAR, 1940. *J. Phys. U.S.S.R.* **3,** 333.
KÜRTI and SIMON, 1938. *Nature, Lond.* **142,** 207.
LANDAU, 1941. *Phys. Rev.* **60,** 358.
LINDEMANN, 1932. *Quantum Theory*, vii. (Oxford).
LONDON, F., 1938. *Nature, Lond.*, **141,** 643; 1939. *J. Phys. Chem.* **43,** 49.
LONDON, F. and LONDON H., 1935. *Proc. Roy. Soc.* A, **149,** 71.
LONDON, H., 1939. *Proc. Roy. Soc.* A, **171,** 484.
NERNST, 1926. *The New Heat Theorem* (London: Methuen).
NIX and SHOCKLEY, 1938. *Rev. Mod. Phys.* **10,** 1.
PICKARD and SIMON, 1945. *Proc. Roy. Soc.* A (to be published shortly).
SHOENBERG, 1938. *Superconductivity* (Cambridge).
SIMON, 1926. *Ber. Akad. Wiss. Berlin,* **33,** 477; 1930. *Erg. Exakt. Naturw.* **9,** 222; 1934. *Nature, Lond.*, **133,** 529.
TELLER and WHEELER, 1938. *Phys. Rev.* **53,** 778.
TISZA, 1938. *C.R. Acad. Sci., Paris,* **207,** 1035, 1186.

# A DOUBLE-REFRACTION METHOD OF DETECTING TURBULENCE IN LIQUIDS

By A. M. BINNIE

*Communicated by E. J. Bowen, F.R.S.* ; *MS. received* 17 *March* 1945

*ABSTRACT.* The doubly-refractive properties of a weak solution of benzopurpurin were used in a study of the onset of turbulence when the liquid flowed through a long horizontal glass pipe. A beam of polarized light was arranged to traverse a diameter of the pipe remote from the ends, the nicols being set to give extinction when the liquid was stationary. The emergent light fell on a photo-cell, which was connected through an amplifier to a cathode-ray tube.

With unidirectional flow, photographs of the trace on the screen showed a steady straight line when the velocity was small. Disturbances appeared at a Reynolds number $R$ of about 1970, and complete turbulence was established when $R$ attained a value of 2900.

With forced oscillating flow, the trace at low velocities was a sinusoidal curve, on which, at a critical Reynolds number, small superposed ripples were observed. Even when the motion was greatly increased the intricate traces recurred perfectly, but this phase finally broke up into complete turbulence as soon as the amplitude of the surface movement in the bottles at the ends of the pipe reached a certain limit.

## §1. INTRODUCTION

CERTAIN colloidal solutions exhibit the property of streaming double refraction. Most of the quantitative experiments which have been carried out on this phenomenon have been concerned with the problem of the shapes of the particles in solution, and they have been performed in the well-known concentric cylinder apparatus. More recently, however, this property has been used by Alcock and Sadron (1934) to provide a means of measuring the velocity distribution in various kinds of laminar flow. The properties of many liquids, suitable for the purpose, have been examined by Weller, Middlehurst and Steiner (1942), who gave numerous references to previous work. Andrade and Lewis (1926), who used rotating cylinders, remarked that the onset of turbulence was favoured by the presence of colloidal particles, but apart from a brief note by Hauser and Dewey (1939), no trace has been found of any attempt to use the method deliberately for detecting turbulence. However, the underlying principle appeared promising for the purpose and worth detailed attention. Accordingly both unidirectional and forced oscillating flow in a circular pipe were examined by this means at the Engineering Laboratory, Oxford. Many investigators * agree that, in the former, laminar flow breaks down at a Reynolds number of about 2100, calculated with the pipe diameter and with the mean velocity of flow. On the other hand, no previous information is forthcoming concerning forced oscillating flow.

---

* Cf. Goldstein (1938).

## §2. THE WORKING LIQUID

The best known colloidal solutions possessing the required properties are vanadium pentoxide ($V_2O_5$), the red dye benzopurpurin 4B and certain clay suspensions. The first two contain filamentous particles; when stationary, the filaments are oriented at random owing to thermal agitation, but when in slow motion they cause the liquid to exhibit doubly refractive properties if viewed at right angles to the direction of flow. It appears from a review by Edsall (1942) of the present state of knowledge that there is a general measure of agreement amongst physical chemists concerning the nature of this effect. In steady laminar flow, the filaments as they are carried with the stream rotate about axes perpendicular to the direction of streaming, but they do so with an angular velocity which is a minimum when they lie along the stream. Thus at any instant there will be more particles in this position than in any other. This explanation also shows why these solutions, unless highly diluted, exhibit non-Newtonian effects, such as the distortion, examined by Lawrence (1935), of the parabolic velocity distribution in laminar unidirectional flow in a pipe. For it is clear that the filaments, as they whirl along, must be comparatively far apart if no interaction is to occur.

Thus for the present purpose it was necessary to select a solution sufficiently weak to show no appreciable anomalous properties, yet strong enough to exhibit measurable optical effects. After some preliminary tests, 0·25% benzopurpurin solution was chosen because it combines ease of preparation in quantity with a comparatively powerful double-refraction effect. The solution was prepared 2 l. at a time; 0·5 gm. of the powder was ground up in a mortar with a little distilled water, and the volume was made up to 0·5 l. After a slight warming the powder was completely dissolved, giving a clear red solution which was then diluted to 2 l. The alternative procedure of adding the powder to the full volume of water was found to be ineffective; the powder then refused to dissolve even on boiling. A solution of 80 gm. of potassium sulphate in 900 cc. of distilled water was then made, and 200 cc. of this was added to each 2 l. of benzopurpurin to form the working liquid. The salt solution causes the dye molecules to aggregate to particles of a size suitable for the double-refraction phenomenon, but larger quantities throw out the dye as a flocculent precipitate. The aggregation is caused by the electrical effects of the monovalent kation; the presence of kations of higher valency, derived, for example, by corrosion, must be guarded against in view of their powerful coagulating effect. Certain parts of the apparatus had to be made of metal, therefore the sulphate was used in preference to the more corrosive chloride.

would be
0.025%

One disadvantage of benzopurpurin as a working liquid is the change of its properties with time. For a few hours after mixing it was found to be inactive. On the first day after mixing its optical power was sufficiently strong for the experiments to proceed, but this increased too much during the day's work for the results to be entirely satisfactory. To avoid this difficulty, observations were then confined to the third and to the seventh days, over which the rate of development was slow. The growth of the particles depends upon the temperature, but this could not be kept constant; hence the optical properties of the liquid on the two test days was not the same in successive sets of experiments. The liquid remained active for periods longer than a week, but it gradually became cloudy and therefore unsuitable.

When it was desired to interpret the observations in terms of Reynolds numbers, considerable difficulty was experienced because the kinematic viscosity of the working fluid was not free from uncertainty. Mark (1940), who described a benzopurpurin solution as "elastic", explained that its viscosity is maximum at very low shearing stresses, diminishing in an exponential manner to a steady Newtonian value as the shearing stress is increased. These effects were discovered in the working fluid, although it was so dilute. Experiments were made with a viscometer of the Couette type, in which the diameter of the outer rotating cylindrical surface was $1\frac{3}{8}$ in. and that of the inner cylinder (hung on a torsion wire) was $\frac{5}{8}$ in. At a mean velocity gradient of 3 sec.$^{-1}$ the viscosity of the three-day-old fluid was found to be 36% in excess of that of water, while at 9 sec.$^{-1}$, the limit of the apparatus, the corresponding value fell to 21%. At a very high velocity gradient, estimated at 800 sec.$^{-1}$, observations with an Ostwald (capillary) viscometer gave the excess of kinematic viscosity as only 3%. The experiments to be described lay in a comparatively small intermediate range, and the onset of turbulence in unidirectional flow was at a mean velocity gradient of about $17\frac{1}{2}$ sec.$^{-1}$. The density of the working fluid was $\frac{1}{2}$% greater than that of water. On the assumption that the excess of kinematic viscosity diminished exponentially with the mean velocity gradient, logarithmic plotting of the above results showed that at $17\frac{1}{2}$ sec.$^{-1}$ the excess was about 9%, and, as no sensible change could be detected after seven days, this figure has been used in all the calculations.

It should be borne in mind that the viscosity of these anomalous solutions is known to depend partly on the previous mechanical history of the solution; vigorous treatment breaks up temporary linkages of the particles and for a time reduces the viscosity. Now in the unidirectional flow experiments the working liquid passed continuously through a centrifugal pump, and the oscillating flow apparatus was operated as violently as possible for a few minutes at the commencement of each series of tests. Therefore viscosities deduced from measurements made under the quiet conditions in the Couette viscometer were probably somewhat higher than those obtaining in the experiments. Support for this view is provided by the unidirectional flow tests described in §4. There it was found that the critical velocity occurred at a Reynolds number of about 1970, which is slightly lower than that established by other methods.

## §3. THE OPTICAL APPARATUS

The horizontal glass pipe, through which the liquid passed, was 15 ft. long. It was made up of three 5-ft. lengths accurately fused together, the mean internal diameter of the central length being $0{\cdot}94_3$ in. At the central section, distant 95 diameters from the ends, the flow was examined by means of a photo-electric cell. The source of light was a 125 w. tungsten arc lamp (pointolite), the horizontal beam from which passed successively through a convex lens (to make the beam parallel), a cell containing a solution of ferrous and copper sulphates (to absorb injurious heat rays), an adjustable stop set at about $^{7}/_{32}$ in. diameter, the first nicol or polarizer, a diameter of the pipe, and finally the second nicol or analyser. The axes of the nicols were fixed at 90° to each other, both being at 45° to the axis of the pipe. In order to preserve the approximately parallel nature of the beam,a rectangular glass cell full of distilled water was fitted round the section of the pipe which was under observation.

When the liquid was stationary, with the dye filaments oriented at random, the only light which emerged was that due to lack of parallelism in the beam, causing unequal reflection losses of the components of the plane-polarized light along and across the tube, with consequent imperfect extinction by the nicols. But, on putting the liquid into motion, the double-refraction effect caused the emergent beam to increase in strength. It fell on a photo-electric electron multiplier, consisting of a caesium photocell whose primary emission was magnified about $2{\cdot}5 \times 10^4$ by seven stages of secondary photo-emission controlled by crossed electromagnetic and electrostatic fields. It is free from inertia effects in rapidly varying light. The output of the multiplier was led through a single stage of valve amplification, having a voltage gain of about 20, to the Y plates of a gas-focused cathode ray tube, the screen of which had a long after-glow. Thus, even at the very low velocities of sweep which were used, it was possible to compare the traces of successive cycles. The time base was of orthodox type, a condenser being charged through a pentode valve and discharged by contacts mechanically operated by the piston rod of the reciprocating pump described in §5. In order to secure the necessary stability, all potentials were supplied by dry batteries and accumulators. With a Leica camera, 56 photographs of unidirectional flow were obtained and 123 of oscillating flow ; those reproduced in figures 2, 5 and 6 were taken at approximately half gain, which was found to be the most suitable ratio because at full gain, in spite of many precautions, the trace was disturbed by stray fields. The zero level of the trace was indicated by projecting the images of spots or arrows on to the screen.

In order that a satisfactory comparison might be made between the two types of flow, no adjustment was made to the glass tube or to the optical apparatus in the course of the work detailed in the following pages. Owing to the difficulty of obtaining a circulating pump, the experiments on oscillating flow were performed first. Here, however, the work will be described in the reverse but logical order.

## §4. THE BREAKDOWN OF UNIDIRECTIONAL LAMINAR FLOW

As shown diagrammatically in figure 1, the inlet end of the glass pipe was joined by hose and a short length of 1-in. stainless-steel pipe to an aspirator bottle A. The outlet end was similarly connected to glass tubing leading to the large bottle B ; this served as a reservoir from which the pump C drew its supply. The pump was of the centrifugal type, driven by an electric motor; its interior was chromium-plated and, to give additional protection, the plating was coated with bakelite varnish. The delivery from the pump passed either to the bottle A through glass tubing and an 11-ft. length of 1-in. stainless-steel pipe or back to the bottle B through a by-pass. Thus the flow through the glass pipe was turbulent at entry ; it was effected by gravity only, and any pulsations developed in the pump could not reach the section under observation. The flow was controlled by means of pinch-cocks on the by-pass and near the outlet end of the glass pipe, and also by altering the speed of the pump. No difficulty was experienced in maintaining constant conditions. The apparatus required a charge of about 8 l. of working fluid.

To measure the flow, a stainless-steel orifice-plate was inserted at D, and open glass stand-pipes fixed to a vertical scale were connected to both sides of it. In view of the limited supply of working fluid, it was thought preferable to use tap-water when determining the coefficients of discharge of the orifice at various rates of flow. In order that the observers should not be biased in any way, this calibration was not done until all the photographs had been obtained.

Observation revealed, and the camera recorded (figure 2), that at low velocities the trace on the screen was a horizontal straight line, which moved upwards further from the zero level when the velocity was increased. The zero with the tungsten arc on and the liquid stationary is indicated on the photographs by two spots; when the light was turned off, the trace descended only by an amount approximately equal to its own breadth. The sweep was set to operate at the constant speed of 20 c./min. At higher velocities of flow, momentary oscillations occurred, evidently due to the "flashes" described by Reynolds. As the velocity was still further increased, the disturbances became more numerous and the frequency of the ripples went up. Finally a velocity was obtained when the trace was always in a state of agitation.

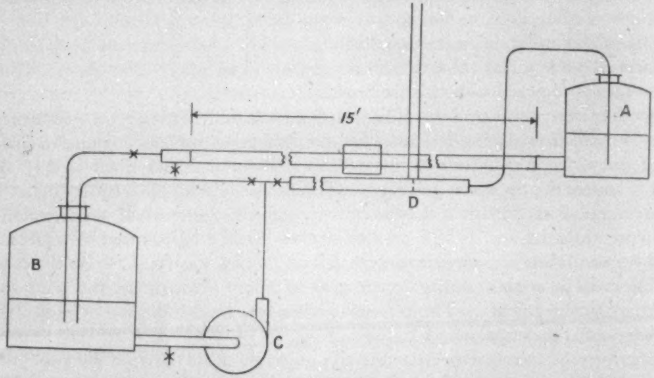
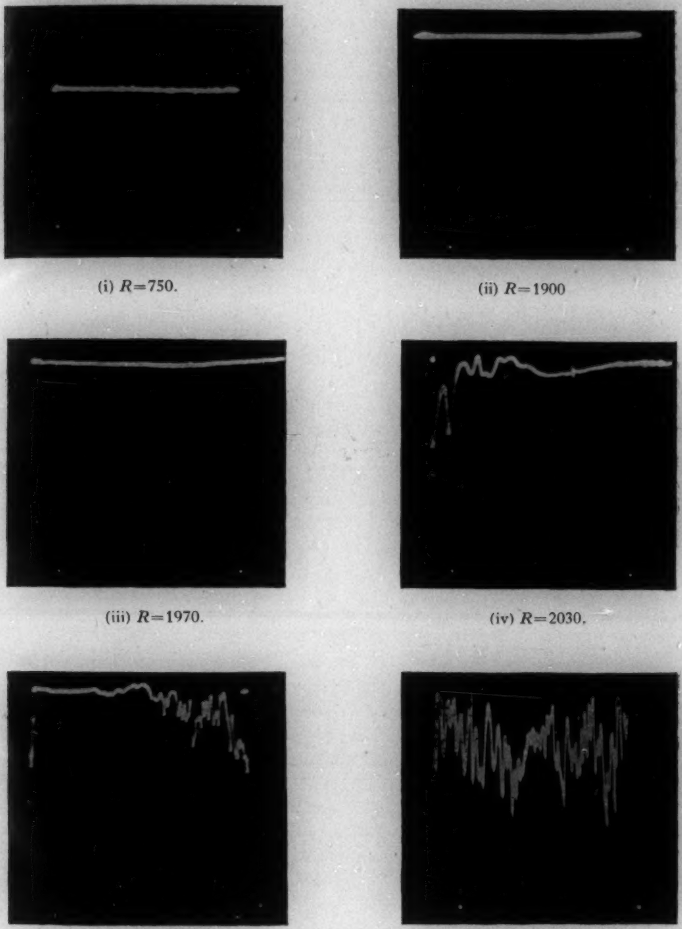


Figure 1. Elevation of the unidirectional flow apparatus.

With the three-day-old liquid, photographs were secured of complete steadiness at a Reynolds number $R = 1970$, and of disturbances at $R = 2270$. But a closer "bracket" was obtained after seven days, therefore a selection of photographs from this test is reproduced in figure 2, where (i) and (ii) show the steady straight lines at Reynolds numbers of 750 and 1900. At $R = 1970$ there were two "flashes" in five minutes, but attempts to photograph them were unsuccessful; however, a picture (iii) taken during the intervening periods shows that the steady



(i) $R=750$.

(ii) $R=1900$

(iii) $R=1970$.

(iv) $R=2030$.

(v) $R=2620$.

(vi) $R=2910$.

Figure 2. Breakdown of laminar unidirectional flow.

line was slightly bowed. At $R=2030$ the disturbances were marked, a period of rippling and its termination being shown on (iv). At higher velocities the quiet periods were shorter and the rippling more violent, although even at $R=2620$ (shown in v) the bursts of turbulence did not often last longer than two cycles. In this phase the trace in an inactive period was seldom perfectly horizontal, and the approach of a turbulent spell could often be detected by a gentle fall of the line. At $R=2910$ (vi) really quiet intervals never occured, and turbulence was completely established, while at still higher velocities the amplitude and frequency of the ripples further increased.

These results show at once that the apparatus gives much the same value of the critical Reynolds number as that obtained by earlier methods, but further information may be obtained from the photographs. Figure 3 gives the distance of the straight traces from the zero with no flow plotted against $R$. Little reliance can be placed on points beyond the critical value because, as explained above, the line in a quiet period was not often truly horizontal, and these measurements are therefore dubious. Confining attention to laminar flow, we see that in the three-day tests the deflexion was linearly dependent on $R$, and this suggests little orientation of the particles at the very low tangential stresses set up in such a wide pipe. Had the orientation been complete, the curve would have become horizontal ; in the seven-day tests it did in fact tend to do so, but the onset of turbulence cut off the finish of the process. These ideas are confirmed by the observation that, when the flow was stopped, the deflexion of the trace did not immediately fall to zero, from which it follows that the particles cannot have been very small. Thus the low degree of orientation must be attributed either to lack of anisometry in the particles, which seems improbable in view of their known filamentous shape, or to the small tangential stresses.

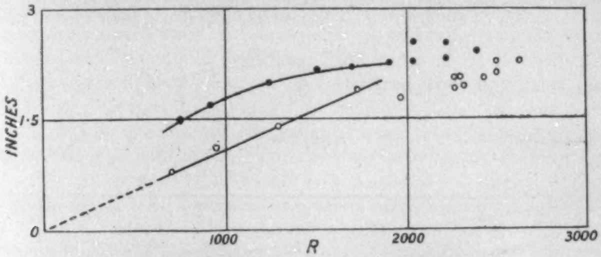


Figure 3. Deflexion of trace in unidirectional flow.
○ After 3 days. ● After 7 days.

The investigation having for its main object the detection of turbulence, no attempt was made to analyse the turbulent vibrations, but it may be remarked that the laminar flow evidently broke up into vortices not much smaller than the cross-section of the light beam used. Only if the vortices were considerably less than this would the double refraction be reduced to zero, and this was not observed.

## §5. THE BREAKDOWN OF OSCILLATING LAMINAR FLOW

It is desirable first to examine theoretically the nature of forced oscillating laminar flow in a pipe—a problem which has been attacked by Southwell and his collaborators, e.g. Christopherson, Gemant, Hogg and Southwell (1938). They considered a U-tube (composed of a horizontal pipe connected at each end to a bottle) in which the motion was maintained by a pulsating air pressure. Being concerned principally with establishing the basis of their method of viscometry, they concentrated their attention on the relation between the amplitudes of the applied pressure and of the surface movement in the bottles. Here, however, we are interested in the distribution of horizontal displacements over a cross-section of the pipe which results from a measured amplitude of the surface movement in the bottles. Accordingly Southwell's analysis has been extended. The details are relegated to the Appendix, but the results are shown in figure 4. This is a vector diagram drawn for the motion in the 15-ft. pipe of water at 15° c. at 8 and 20 c./min., these frequencies being the slowest and fastest used in the experiments. The direction of the line Ox represents the amplitude in the bottles. The curve is graduated in terms of the ratio of the radius $r$ under consideration to the radius $a$ of the pipe, and the magnitude and phase angle of the horizontal displacement are given by the straight line drawn from O to the curve. The distribution of these displacements has a kind of inverse analogy to the well-known skin effect which occurs when high-frequency alternating current passes through a solid circular conductor. It will be seen that at 8 c./min. the liquid within a radius of about $0{\cdot}4\,a$ oscillates with

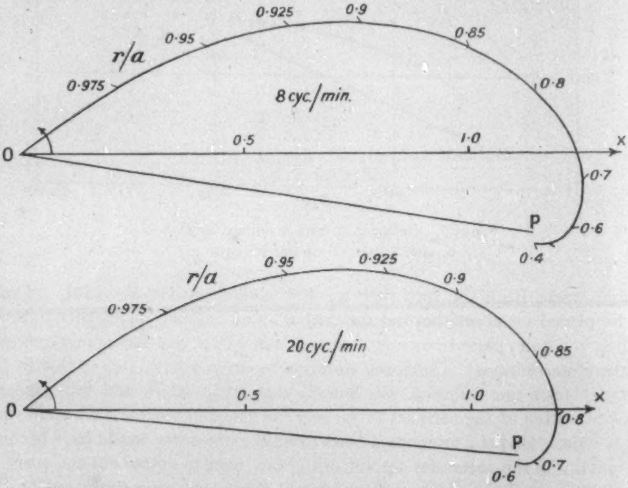


Figure 4. Theoretical distribution of displacement in laminar oscillating flow.

little relative movement. At 20 c./min. this radius is increased to about 0·6 $a$; it is only at frequencies beyond the scope of the experiments that almost the whole of the liquid oscillates virtually as a solid plug with large shearing stresses close to the wall. It should be noted that the ratio of the horizontal displacements in the pipe to the amplitude in the bottles is independent of the latter when the other variables are unchanged.

In the first set of experiments the glass pipe was connected at each end to an aspirator bottle of diameter 5·6 in. In one bottle, off which the top was cut, the movement was measured with a stainless-steel hook gauge, surface disturbances being eliminated by perforated bakelite baffles. The other bottle was joined by rubber tubing to a single-cylinder reciprocating pump of diameter 3 in. and stroke 1⅝ in., from which the valves were removed. The piston was actuated by a Scotch crank driven through reduction gearing by an electric motor, so that its motion was truly simple harmonic. It was found that if, on starting, the rubber tubing was connected when the piston was at the middle of its stroke, no creep in the excursions of the liquid took place. The amplitude of the motion of the liquid was controlled with four bottles (serving as capacities), joined singly to the connexion between the pump and the aspirator bottle; they could be turned on or off at will, and the effective volume of one could be finely adjusted by partially filling it with water. To satisfy the requirement that the pressure imposed on the working liquid should vary in a sinusoidal manner, the pressure changes in the air were kept small and nearly isothermal by running the pump only at very low speeds. The apparatus required about 5 l. of working liquid.

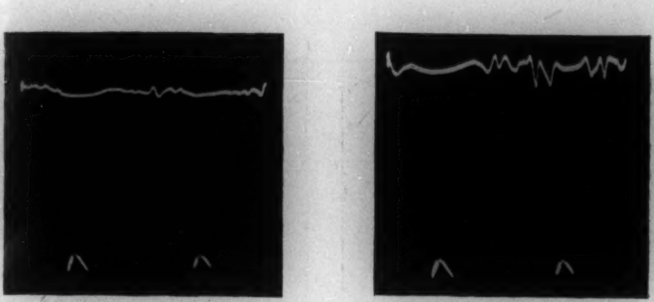
The method of test was to take a series of photographs at constant pump speed, the amplitude of the movement in the aspirator bottles being varied by means of the capacities. The surprising fact at once emerged that over a considerable range the trace at each amplitude, although considerably agitated, recurred exactly for an indefinite length of time. The break-up of this phase at 13 c./min. is illustrated by the series shown in figure 5, in which are indicated the travels measured in the aspirator bottle and the Reynolds numbers. The latter will now be formed with the maximum velocity of the motion, with the simplifying assumption that the velocity was uniform over the cross-section. The zero at no flow is here indicated by the tips of the arrows. The first photograph shows recurring flow at a travel of 0·55 cm.; the ends of the stroke correspond to the undisturbed portions of the trace, which, as in the unidirectional experiments, did not instantly descend to zero when the velocity became very small. At 0·80 cm., (ii), the trace still recurred. The next photograph records two consecutive cycles, the only difference which can be detected between them being the short-circuit of the prominent inverted W on the right. Again, the ends of the stroke may be distinguished by the quieter motion there, which, being slower, caused a thickening of the trace. Finally, at 1·61 cm., (iv), the diagrams did not recur, although the differences between them was not marked, but at still greater travels the motion of the trace was wildly irregular. These results indicate that under certain conditions a recurring system of vortices was formed, which was destroyed when the motion became too violent. The available information concerning the region of breakdown is summarized in table 1, in which the results of the one-day test (mentioned in § 2) have been included.

Table 1. The breakdown of recurring turbulent flow

| Age of fluid | c./min. | Travel (cm.) | $R$ | |
|---|---|---|---|---|
| One day | 20 | 0·97 | 6360 | Recurring |
| ,, | ,, | 1·30 | 8480 | Almost recurring |
| ,, | ,, | 1·41 | 9220 | Not recurring |
| ,, | 16½ | 1·24 | 6840 | Recurring |
| ,, | ,, | 1·52 | 8360 | Not recurring |
| ,, | 13 | 1·03 | 4610 | Recurring |
| ,, | ,, | 1·39 | 5880 | Almost recurring |
| ,, | ,, | 1·51 | 6750 | Not recurring |
| Three days | 20 | 0·98 | 6500 | Recurring |
| ,, | ,, | 1·32 | 8800 | Not recurring |
| ,, | 13 | 1·08 | 4840 | Recurring |
| ,, | ,, | 1·39 | 6190 | Not recurring |
| Seven days | 20 | 0·95 | 5810 | Recurring |
| ,, | ,, | 1·32 | 8050 | Not recurring |
| ,, | 13 | 0·80 | 3480 | Recurring. |
| ,, | ,, | 1·10 | 4790 | Almost recurring |
| ,, | ,, | 1·25 | 5460 | Almost recurring |
| ,, | ,, | 1·61 | 6990 | Not recurring |

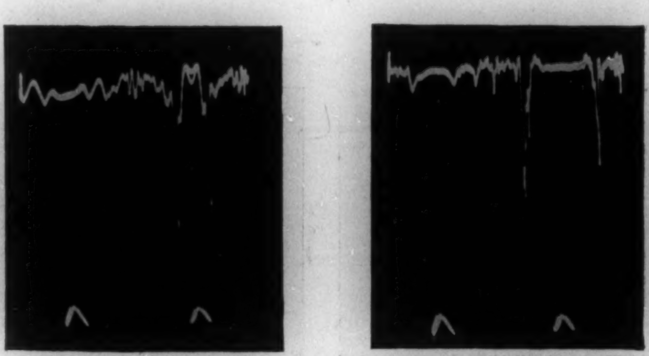
No correlation can be discovered among the values of $R$, but there are indications that it was the travel which was the determining factor. Allowing for the difficulty of estimating when recurrence was perfect, it may be suggested that a travel of about 1·25 cm. was the criterion of breakdown.

Attention was then concentrated on the first signs of turbulence. With a fresh charge of liquid, the trace at very slow flows was seen to be a smooth sinusoidal curve indicating laminar conditions. As the amplitude was increased, small ripples appeared at first, but later the harmonic nature of the picture was broken up into the form indicated in figure 5 (i). These experiments were found difficult and unsatisfactory because under some conditions the travel did not remain steady. Consequently suspicion fell on the apparatus used to maintain the motion. It has been pointed out by den Hartog (1940) that, if a forced oscillation is maintained with a spring of varying stiffness as coupling, the resulting motion may be unstable. In the experiments, the damping was considerable owing to the destruction of kinetic energy when the liquid entered an aspirator bottle ; free oscillations were in fact nearly dead-beat. Thus the power required to drive the apparatus, and consequently the compression and varying elasticity of the intermediate air, seemed unnecessarily large. Moreover, at 20 c./min. the travel required for laminar flow was too small to be measurable with reasonable percentage accuracy ; and when the frequency was halved in order that a wide



(i) Travel 0·55 cm., Max. $R=2400$.

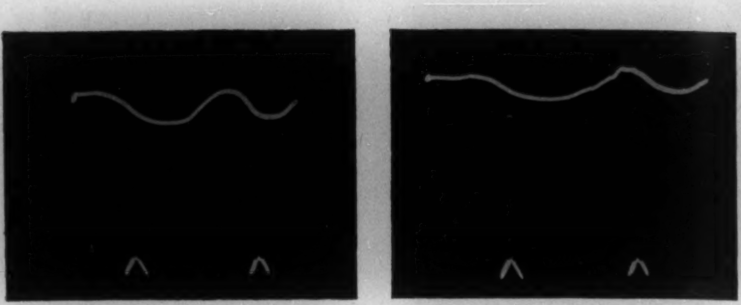
(ii) Travel 0·80 cm., Max. $R=3480$.



(iii) Travel 1·25 cm., Max. $R=5460$.

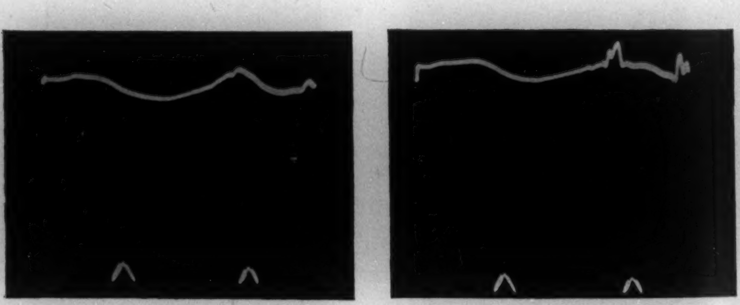
(iv) Travel 1·61 cm., Max. $R=6990$.

Figure 5. Breakdown of recurring oscillating flow.



(i) Travel 4·39 cm., Max. $R=730$.

(ii) Travel 5·83 cm., Max. $R=980$



(iii) Travel 6·25 cm., Max. $R=1050$.

(iv) Travel 8·22 cm., Max. $R=1360$.

Figure 6. Breakdown of laminar oscillating flow.

range of conditions might be investigated, it approached so closely to the natural period that sufficiently small travels were unobtainable. Accordingly the aspirator bottles were replaced by vertical glass tubes, 19 in. long and only $1\frac{5}{8}$ in. in diameter, the lower portions of which were bent through an easy right angle. A carboy was substituted for the four capacities, and the resistance to the motion of the air was reduced by increasing the diameter of the connecting tubing from $\frac{1}{4}$ to $\frac{1}{2}$ in. The travel was measured with a horizontal telescope mounted on a cathetometer.

As expected, these modifications greatly reduced the damping, and tests were made at frequencies both above and below the natural period, but the difficulties mentioned above did not disappear. A test after three days at 20 c./min. showed, after a considerable wait, a smooth sinusoidal curve at $R=890$. At the next stage, a recurring trace with slight superposed ripples was obtained after allowing it a long time to settle, but the travel never became absolutely steady: its mean was 2·51 cm. (corresponding to which $R$ is 1050), but the excursions varied by about $\frac{1}{2}$ mm. at each end. These irregularities in the travel continued up to $R=1320$, accompanied by rippling of increasing violence, but at and beyond $R=1480$ no trouble was experienced in quickly obtaining both recurring traces and steady travels. Other observations made it clear that it was the Reynolds number, not the travel, which was the criterion of the onset of turbulence and of irregular travel. This conclusion is supported by the illustrations displayed in figure 6, which were taken with fluid seven days old at a speed of only $8\frac{1}{2}$ c./min. Although the sweep unfortunately gave trouble at this point, (i) shows that at $R=730$ the trace was smooth. The displacement of the trace being greater than in the three-day tests described above, at $R=800$ and 920 it was just possible to detect faint ripples, but they are too small to be seen on the necessarily small scale to which the photographs are reproduced here. At $R=980$, (ii), the rippling was more marked; at $R=1050$, (iii), however, the difficulty of obtaining steady travel again commenced, although the trace recurred. The irregularities were at their height at $R=1240$: here the trace apparently settled down quickly, but then a spasm occurred and the trace slowly changed its form, not settling down to its original shape until 20 min. had elapsed. The recurring diagram, (iv), at $R=1360$ was taken after 15 min., the last 5 min. being steady, but at and beyond $R=1610$ no further trouble was encountered.

The results suggest a critical value of $R$, not very sharply marked, at about 800. The existence of a zone of irregular travel at slightly greater values of $R$ may be attributed, not to the drive, but to the formation of small eddies, principally at the ends of the pipe. These would erratically increase the small viscous resistance to motion, and, when sufficiently large, would cause a noticeable variation in the travel. At greater amplitudes they would become more numerous and possess an almost constant mean value, and under these conditions a steady travel would again be obtained. It thus appears that the onset of turbulence can be detected, although in a less sensitive manner, simply by examining the uniformity of the travel.

## §6. CONCLUSIONS

(i) Considerable difficulty was encountered in securing a satisfactory compromise between the conflicting requirements of the various parts of the apparatus. A wide pipe and a narrow beam of light were employed so that the lens effect of the former was small. Thus the emergent beam was sensibly parallel, and it was possible to obtain good extinction with the analyser. The disadvantage of the wide pipe was a very low critical velocity, occurring at such small shearing stresses that a slight anomaly in the viscosity of the working liquid, with its attendant distortion of the velocity distribution, was present. If, to avoid these effects, the working liquid had been a weaker solution, even more sensitive electrical apparatus would have been necessary.

(ii) With unidirectional laminar flow, the trace was a horizontal straight line. At a Reynolds number of 1970, brief disturbances appeared, which became more numerous and violent as the velocity was increased. Thus the indications of the apparatus were in fair accordance with the results obtained by other methods. Completely developed turbulence was shown to set in at a Reynolds number of about 2900.

(iii) With oscillating laminar flow, the trace was a smooth sinusoidal curve. At a certain Reynolds number (derived in an approximate manner) recurring ripples appeared, which suggested the formation of a steady system of vortices. Over a small range of slightly higher Reynolds numbers the motion of the liquid became slightly erratic owing it is thought, to the formation of eddies which irregularly increased the viscous resistance. Under much more vigorous conditions of flow the recurring traces broke up, indicating complete turbulence, and this point was found to depend upon the travel of the liquid.

## ACKNOWLEDGMENTS

The investigation described in the foregoing pages is the result of much cooperative effort. Mr. E. J. Bowen, F.R.S., suggested to me that the double-refraction effect was worth studying for the purpose of detecting turbulence, and he assisted in innumerable ways at every stage of the research. Mr. J. G. G. Hempson assembled the valve amplifier and the cathode-ray tube to his own design. Mr. V. Belfield carried out the whole of the photographic work, Dr. B. Lambert undertook the difficult task of jointing the 15-ft. glass tube, and Mr. W. M. Aitken assisted with the observations. To all these gentlemen I desire to express my grateful thanks. I am also greatly indebted to Messrs. Ricardo & Co., Ltd., for the loan of the reciprocating pump.

## APPENDIX

### *Oscillatory horizontal displacements in the pipe*

An expression is required giving the distribution of horizontal displacements over the cross-section of the pipe in terms of the amplitude of the surface movement in the bottles.

We adopt the notation used by Southwell and his collaborators (loc. cit.), and refer to their results by inserting the numbering of their equations at the left-hand margin of the page. Let $p$ be the externally applied pressure, $x$ the displacement in the bottles, $v$ and $y$ the horizontal velocity and displacement in the pipe at radius $r$. The uniform cross-section of the bottles is denoted by $A$, $a$ is the radius of the pipe, $\mu$ and $\rho$ are the viscosity and density of the liquid, $2h$ is the length of the pipe, and $g$ is the acceleration due to gravity.

We assume that

(4)
$$\left.\begin{aligned} p &= Pe^{int}, \\ x &= Xe^{int}, \\ v &= Ve^{int}, \\ y &= Ye^{int}, \end{aligned}\right\} \qquad \ldots\ldots\text{(i)}$$

where $P$, $X$, $V$ and $Y$ are complex, and $V$ and $Y$ are functions of $r$ only. Then it was proved that

(8)
$$V = \frac{1}{k^2}\left(Q - \frac{P}{2h\mu}\right)\left[1 - \frac{J_0(kr)}{J_0(ka)}\right], \qquad \ldots\ldots\text{(ii)}$$

(5) where
$$Q = \frac{2\pi g}{hA}\frac{k^2}{n^2}\int_0^a rV\,dr,$$

and
$$k^2 = -in\rho/\mu.$$

(9) But
$$X = \frac{h\mu}{g\rho}Q,$$

(10) and
$$\left[\frac{g'}{hn^2} - \left\{1 - \frac{2}{ka}\frac{J_1(ka)}{J_0(ka)}\right\}^{-1}\right]X = \frac{g'}{g}\frac{P}{2\rho hn^2},$$

(7) where
$$g' = \frac{\pi a^2}{A}g.$$

Hence, on integrating (ii), we obtain, after some reduction,

$$\frac{Y}{X}\frac{g'}{g} = \left\{1 - \frac{2}{ka}\frac{J_1(ka)}{J_0(ka)}\right\}^{-1}\left(1 - \frac{J_0(kr)}{J_0(ka)}\right) + E, \qquad \ldots\ldots\text{(iii)}$$

where $E$ is a function of $r$ only. It is evidently zero when a steady state of oscillation has been attained. Now $k^2$ is a pure imaginary, hence we introduce a quantity $z$ given by

(12)
$$z^4 = -k^4a^4.$$

Then (iii) may be written

$$\frac{Y}{X}\frac{g'}{g} = \left\{1 - \frac{2}{iz\sqrt{i}}\frac{J_1(iz\sqrt{i})}{J_0(iz\sqrt{i})}\right\}^{-1}\left(1 - \frac{J_0(i\omega\sqrt{i})}{J_0(iz\sqrt{i})}\right). \qquad \ldots\ldots\text{(iv)}$$

Here $z$ and $\omega$ are real, and they are given by

$$z = a\left(\frac{n\rho}{\mu}\right)^{\frac{1}{2}}, \quad \omega = r\left(\frac{n\rho}{\mu}\right)^{\frac{1}{2}}.$$

The evaluation of the right-hand side of (iv) may be performed in two parts. The first term reduces to

$$(20)\quad \left\{\begin{array}{l} \qquad\qquad\qquad C(z)-iD(z), \\ \text{where}\ C(z)=1+\dfrac{A(z)-1}{\{A(z)-1\}^2+\{B(z)\}^2},\quad D(z)=\dfrac{B(z)}{\{A(z)-1\}^2+\{B(z)\}^2}; \\ \text{and}\quad A(z)=\dfrac{z}{2}\dfrac{W(z)}{V(z)},\quad B(z)=\dfrac{z}{2}\dfrac{Z(z)}{V(z)}. \end{array}\right\} \qquad \ldots\ldots(\text{v})$$

The modulus of this vector is $[\{C(z)\}^2+\{D(z)\}^2]^{\frac{1}{2}}$ and its phase angle $\theta$ is given by $\tan\theta=-D(z)/C(z)$. Numerical values of these expressions can be obtained with the aid of the tables given by Russell (1914). The second term in (iv), which gives the displacement distribution over the cross-section, may be expeditiously dealt with by writing it in the form

$$1-\frac{M_0(\omega)}{M_0(z)}\,e^{i\{\theta_0(\omega)-\theta_0(z)\}},$$

and making use of McLachlan's tables (1934).

The diagrams of figure 4 can then be constructed. Taking OX to represent the direction of the vector $X$, we set off OP of length $[\{C(z)\}^2+\{D(z)\}^2]^{\frac{1}{2}}$ at the phase angle $\theta$. For various values of $r/a$, rays from P are drawn of length $[\{C(z)\}^2+\{D(z)\}^2]^{\frac{1}{2}}\,M_0(\omega)/M_0(z)$, the angles between them and OP being $\{\theta_0(\omega)-\theta_0(z)\}$. Lastly, the curve joining the ends of the rays is drawn in.

## REFERENCES

Alcock, E. D. and Sadron, C. L., 1935. *Physics*, **6**, 92.

Andrade, E. N. da C. and Lewis, J. W., 1926. *Kolloidzschr*. **38**, 260.

Christopherson, D. G., Gemant, A., Hogg, A. H. A. and Southwell, R. V., 1938. *Proc. Roy. Soc.* A, **168**, 351.

den Hartog, J. P., 1940. *Mechanical Vibrations*, 2nd ed. p. 387 (New York and London: McGraw-Hill).

Edsall, J. T., 1942. *Advances in Colloid Science* (ed. Kraemer, E. O.), vol. i, p. 269 (New York: Interscience Publishers).

Goldstein, S. (ed.), 1938. *Modern Developments in Fluid Dynamics*, p. 319 (Oxford: Clarendon Press).

Hauser, E. A. and Dewey, D. R., 1939. *Industr. Engng. Chem.* **31**, 786.

Lawrence, A. S. C., 1935. *Proc. Roy. Soc.* A, **148**, 59.

Mark, H., 1940. *High Polymers*, vol. ii, fig. 82 (New York: Interscience Publishers).

McLachlan, N. W., 1934. *Bessel Functions for Engineers*, p. 182 (Oxford: Clarendon Press).

Russell, A., 1914. *The Theory of Alternating Currents*, 2nd ed. vol. i, p. 233 (Cambridge: University Press).

Weller, R., Middlehurst, D. J. and Steiner, D. J., 1942. *N.A.C.A. Technical Note*, no. 841.

# EXACT ADDITION FORMULAE FOR THE AXIAL SPHERICAL ABERRATION AND CURVATURE OF FIELD OF AN OPTICAL SYSTEM OF CENTRED SPHERICAL SURFACES

By F. GILBERT BROWN,
Bexley

*Communicated by Prof. L. C. Martin. MS. received* 4 *February* 1945 ;
*in revised form* 14 *May* 1945

*ABSTRACT.* From a pair of axial magnification formulae, exact addition equations are derived in this paper for the axial spherical aberration and curvature of field for central and sagittal rays produced by any number of centred spherical surfaces whatever their individual separations.

For a ray of initial semi-aperture $U_1$ and final semi-aperture $U_q'$ passing through media of refractive indices $n_1\ n_2\ \ldots\ n_q$, the condition for zero axial spherical aberration is

$$\sum_{p=q}^{p=1} n_p u_p c_p (\sin U_p + \sin I_p - \sin I_p' - \sin U_p') = 0,$$

where $c_p$ is the distance of the $p$th intermediate paraxial object point from the centre of curvature of the $p$th surface.

For zero Petzval curvature and an initial semi-angle of field $\theta$, the condition is first given in the usual approximate Petzval form with the addition of a corrective term

$$\sum_{p=q}^{p=1} \frac{n_p' - n_p}{n_p' n_p r_p} \left\{ 1 + \tan^2 \frac{\theta_p}{2} - 2 \frac{n_I'}{n_p} m_p \left( \frac{x_p}{Y_p} - \tan \frac{1}{2}\theta_p \right)^2 \right\} = 0.$$

A more simple alternative form, however, is $\sum_{p=q}^{p=1} \delta_p \tan \frac{1}{2}\theta_p = 0$, where $\delta_p$ is the deviation of the central paraxial ray at the $p$th surface.

An addition equation is also given for the sagittal astigmatism.

---

## § 1. INTRODUCTION

Exact values for the aberrations of the image formed by compound optical systems of centred spherical refracting surfaces are usually obtained by trigonometrical ray-tracing, and the removal of the aberrations, although supplemented by well-known approximate formulae in many cases, thus depends in some degree on a trial-and-error process, in which values for the unknown radii of curvature are first assumed and then modified or adjusted until the selected aberrations are removed.

The object of this paper is to show that so far as the axial spherical aberration, the curvature of field for central rays (Petzval curvature), and to a lesser degree the sagittal astigmatism, are concerned, exact addition formulae can be derived, containing a minimum number of variables, which, when equated to zero, give the direct conditions for the absence of these aberrations.

It is hoped that the publication of these formulae may, in combination with the well-known use of the sine condition for the pre-determination of coma, indicate useful means of estimating in advance the magnitude of the several aberrations in any given system,.and at the same time encourage further development along these lines.

## § 2. THE AXIAL MAGNIFICATION FORMULAE

The formulae developed in this paper are based upon a pair of axial magnification formulae connecting (*a*) the successive intercepts $\Delta l$ and $\Delta l'$ cut off respectively before and after refraction by a pair of paraxial rays, and (*b*) the corresponding intercepts $\Delta L$ and $\Delta L'$, when one of the pair is paraxial and the other of finite semi-aperture $U$.

For infinitesimal lengths $dl$ and $dl'$ it is well known that the relation for paraxial rays is

$$dl' = \frac{n'}{n} m^2 dl,$$

but for intercepts of finite length, this equation is an approximation only.

The exact equation of transfer for a pair of paraxial rays is

$$\Delta l' = \frac{n'}{n} m m_0 \Delta l \quad \text{or} \quad n'u'u_0'\Delta l' = nuu_0\Delta l, \qquad \ldots\ldots(1)$$

where $m$ and $m_0$ are the linear magnifications at the respective extremities of the length $\Delta l'$, $u$ and $u_0$ are the angles between the respective paraxial rays and the axis before refraction, and $u'$ and $u'_0$ are the corresponding angles after refraction. This equation may be deduced from the familiar paraxial formulae

$$ff' = -gg' = -g_0g_0' \quad \text{and} \quad f/g = m f'/g_0 = n'm_0/n,$$

remembering that $\Delta l' = g' - g_0'$, where the distances $g$, $g_0$, $g'$, $g_0'$ are measured from the principal foci $f$ and $f'$ respectively.

It is sometimes convenient to substitute $m$ for $m_0$, by putting $g + \Delta l$ for $g_0$, so that

$$\frac{m}{m_0} = \frac{g+\Delta l}{g} = 1 + \frac{\Delta l}{g} = 1 + \frac{m\Delta l}{f},$$

whence

$$\frac{1}{m_0} = \frac{1}{m} + \frac{\Delta l}{f};$$

similarly

$$\frac{1}{m} = \frac{1}{m_0} - \frac{\Delta l}{f}. \qquad \ldots\ldots(2)$$

Substituting the value of $m_0$ above in equation (1),

$$\Delta l' = \frac{n'}{n} m^2 \Delta l \left(1 - \frac{m_0 \Delta l}{f}\right), \qquad \ldots\ldots(2b)$$

an equation which will be useful later on in the paper.

Next, let one of the rays be of finite aperture $U$ cutting the optical axis in B and B′ before and after refraction at distances $C$ and $C'$ from the centre of curvature of the refracting surface (see figure 1). Let $b$ and $b'$ denote the corresponding intersections of the paraxial ray at distances $c$ and $c'$ from the centre. Finally, let $b_0'$ represent the position of the paraxial image of B, and $c_0'$ its distance from the centre of curvature.

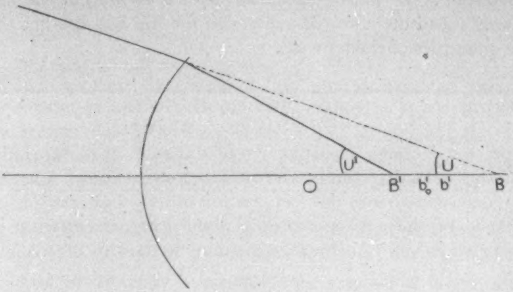


Figure 1.

$OB = C = c_0$.
$OB' = C'$.
$Ob = c$.
$Ob' = c'$, paraxial image of $b$.
$Ob_0' = c_0'$, paraxial image of $B$.

$Bb = \Delta L = \Delta l$.
$B'b' = \Delta L'$.
$b_0'b' = \Delta l'$.
$B'b' = A$ = longitudinal spherical aberration.

Then, using $\Delta L'$ and $\Delta L$ in place of $\Delta l'$ and $\Delta l$ in order to indicate that one of the image-forming rays is now of finite aperture, and denoting the magnifications at b', $b_0'$ and B' by $m$, $m_0$ and $M$ respectively, we have, from equation (1),

$$\Delta L' = \frac{n'}{n} m m_0' \, \Delta L + A,$$

where $A$ is the longitudinal spherical aberration of the image of the point B for semi-aperture $U$.

Substituting $c_0'/C$ for $m_0$ in the preceding equation, we have

$$\begin{aligned} \Delta L' &= \frac{n'}{n} m \frac{c_0'}{C} \, \Delta L + A \\ &= \frac{n'}{n} m \frac{C' + A}{C} \, \Delta L + A \\ &= \frac{n'}{n} m M \Delta L + A \left(1 + \frac{n'}{n} m \frac{\Delta L}{C}\right), \text{ since } \frac{C'}{C} = m_0. \end{aligned}$$

Again, since $\frac{n'}{n} m m_0 \Delta L = \Delta l'$,

$$\frac{n'}{n} m \frac{\Delta L}{C} = \frac{c_0}{c_0'} \frac{\Delta l'}{c}.$$

Hence

$$\Delta L' = \frac{n'}{n} m M \Delta L + \frac{c'}{c_0'} A = \frac{nu \sin U}{n'u' \sin U'} \Delta L + \frac{c'}{c_0'} A. \qquad \ldots\ldots(3)$$

In a system of centred spherical refracting surfaces, the $\Delta L'$ of one surface becomes the $\Delta L$ of the next, and the $U'$ of the one surface the $U$ of the next. Equation (3) may therefore be extended to

$$n_q'u_q' \sin U_q' \Delta L_q' = n_1 u_1 \sin U_1 \Delta L_1 + \sum_{p=q}^{p=1} n_p'u_p' \sin U_p' A_p \frac{c_p'}{c_{0p}'}. \qquad \ldots\ldots(4)$$

If in equation (4) a substitution can be made for the $A_p$ in terms of the known elements of the $p$th surface, we have available an addition equation for spherical aberration. By a happy accident, as will be seen in the next section, we can not only substitute a simple expression for $A_p$, but also get rid of all the troublesome quantities of the type $c'_{0p}$.

## § 3. SPHERICAL ABERRATION OF A RAY OF FINITE APERTURE $U$, PRODUCED BY A SINGLE SPHERICAL SURFACE, SEPARATED BY MEDIA OF REFRACTIVE INDICES $n$ AND $n'$

In the present section, departing from accepted procedure, the radii of curvature of the refracting surfaces are treated as unknown quantities to be subsequently determined, and the formula for spherical aberration is based, in the first instance, on the semi-apertures $U$ and $U'$ before and after refractions. These are selected as known quantities and may be said to determine the *shape* of the system.

It only remains to fix a single linear constant (conveniently represented by paraxial $c'$) to determine the *scale*, after which all other linear quantities are uniquely determined, including the curvature of the surface and the spherical aberration.

By definition we have

$$\frac{A}{C}=\frac{C'-c'}{C}=\frac{n\sin U}{n'\sin U'}-\frac{nu}{n'u'}=\frac{nu}{n'u'}\left(\frac{au'-\sin U'}{\sin U'}\right),$$

if $\sin U=au$.

If $u$ be taken so that $a=1$, the preceding equation becomes

$$A\sin U'=-c'\,(\sin U'-u'),\ \text{since } nuc=n'u'c'.$$

Alternatively this may be written:

$$A\sin U'=c'\,(\sin U+\sin I-\sin I'-\sin U'), \qquad \ldots\ldots(5)$$

for, as $a=1$, $u'=u+i-i'=\sin U+\sin I-\sin I'$, where $i$ and $I$ represent the angles of incidence for the paraxial and marginal ray respectively, and $i'$ and $I'$ the corresponding values for the refracted ray.

The expression within the brackets will henceforth be designated as $\Delta(UU')$. Its value for any given ray of semi-aperture $U$ and $U'$ before and after refraction may be readily determined (by interpolation if necessary) from a set of specially prepared tables * in which the value of $\sin I-\sin I'$ is tabulated for standard values of the deviation $D=U'-U=I-I'$, and for the relative refractive index $n'/n$.

Alternatively $\sin I-\sin I'$ may be replaced by $i-i'=\delta$, so that $\Delta(UU') =\{(u+\delta)-\sin(U+D)\}$, the value of $D$, in terms of $\delta$ being given by the series

$$D=\delta-\tfrac{1}{2}\,\tfrac{1}{3}\frac{n'^3-n^3}{(n'-n)^3}\delta^3+\tfrac{1}{2}\cdot\tfrac{3}{4}\cdot\tfrac{1}{5}\frac{n'^5-n^5}{(n'-n)^5}\delta^5-\tfrac{1}{2}\cdot\tfrac{3}{4}\cdot\tfrac{5}{6}\cdot\tfrac{1}{7}\frac{n'^7-n^7}{(n'-n)^7}\delta^7\ldots+\text{etc.} \qquad \ldots\ldots(6)$$

For if we expand $I$ in terms of $\sin I$, and $I'$ in terms of $\sin I'$, we have

$$I=\sin I-\tfrac{1}{2}\cdot\tfrac{1}{3}\sin^3 I+\tfrac{1}{2}\cdot\tfrac{3}{4}\cdot\tfrac{1}{5}\sin^5 I\ldots\text{etc.},$$

and similarly for $I'$.

---

* The writer prepared some time ago a set of such tables, covering deviations between 0° and 12° and values of log $n'/n$ between 18000 and 22000.

Hence

$$D = I - I' = (i - i') - \tfrac{1}{2}\cdot\tfrac{1}{3}\left(\frac{n'^3 - n^3}{n'^3}\right) i^3 + \text{etc., since } \sin I' = \frac{n}{n'}, \sin I \text{ and } i = \sin I.$$

Then equation (6) follows, since $\delta = i - i' = \dfrac{n' - n}{n'} i$, $i^3 = \dfrac{n'^3}{(n'-n)^3}\delta^3$; $i^5 = \dfrac{n'^5}{(n'-n)^5}\delta^5$; and so on.

The series in equation (6) is rapidly convergent for small values of $\delta$, but for the rapid evaluation of $\Delta(UU')$ it is quicker to pick out from a table of logarithmic sines a pair of values for log sin $I$ and log sin $I'$, which differ by log $n'/n$, while $I$ and $I'$ differ by the required deviation $U' - U$.

## §4. AXIAL SPHERICAL ABERRATION OF THE MARGINAL RAY AFTER REFRACTION BY *q*-CENTRED SPHERICAL SURFACES

We can now substitute in equation (4) the value of $A$ found in equation (5) above, noting that the $c_0'$ in the former equation corresponds to the $c'$ in equation (5) for each successive surface.

Hence

$$n_q' u_q' \sin U_q' A_q = n_1 u_1 \sin U_1 A_1 + \sum_{p=q}^{p=1} n_p' u_p' c_p' \Delta(U_p U_p'). \qquad \ldots\ldots(7)$$

In the absence of initial spherical aberration, the condition that the final ray $U_q'$ shall be free from spherical aberration is

$$\sum_{p=q}^{p=1} n_p' u_p' c_p' \Delta(U_p U_p') = 0. \qquad \ldots\ldots(8)$$

Thus the general procedure in building up a system free from axial spherical aberration is first to define the path of the marginal ray through the system by specifying the $U'$ of each successive surface (the $U_p$ of the $p$th surface being identical with the $U'_{p-1}$ of the preceding). These deviations having been chosen in accordance with the requirements of achromatism and the desired angular magnification of the system, factors are then selected representing the partial product $n_p' u_p' c_p'$ for each surface, in such a way as to make the algebraic sum on the left-hand side of equation (7) zero. This selection must of course be conditioned by other practical requirements of the system, such as the lens thicknesses, the separations of the individual lenses, and the intersection heights, as well as by the need to remove other aberrations.

## §5. CONTROL OF COMA

Control of coma will normally be achieved by the well-known application of the sine condition. With an initial $u_1 = \sin U$, it will merely be necessary to see that final $u_q' = \sin U_q'$.

In the early stages of laying out the system, however, $u_q'$ has not been determined, its exact value being dependent on the linear quantity $c$, and we only know the product $uc$. If, however, we write

$$k_p = n_p u_p c_p \quad \text{and} \quad K_p = \frac{\Sigma k_p \Delta(UU')_p}{\sin U_p'},$$

then the ratio of the magnification $M_p$ of the marginal ray to the magnification $m_p$ of the paraxial ray is

$$\frac{k_p(k_1+K_1)(k_2+K_2)(k_3+K_3)\,\ldots\ldots\ldots\,(k_{p-1}+K_{p-1})}{k_1(k_2+K_1)(k_3+K_2)(k_4+K_3)\,\ldots\ldots\ldots\,(k_p+K_{p-1})}, \quad \ldots\ldots(9)$$

which must be equal to unity if the final coma is to be zero.

For
$$M/m=\frac{\left(1+\frac{A_1}{C_1'}\right)\left(1+\frac{A_2}{C_2'}\right)\left(1+\frac{A_3}{C_3'}\right)\ldots\ldots\ldots\left(1+\frac{A_p}{C_p'}\right)}{\left(1+\frac{A_0}{C_1}\right)\left(1+\frac{A_1}{C_2}\right)\left(1+\frac{A_2}{C_3}\right)\ldots\ldots\ldots\left(1+\frac{A_{p-1}}{C_p}\right)}$$
$$=\frac{\left(1+\frac{K_1}{k_1}\right)\left(1+\frac{K_2}{k_2}\right)\left(1+\frac{K_3}{k_3}\right)\ldots\ldots\ldots\left(1+\frac{K_p}{k_p}\right)}{\left(1+\frac{K_0}{k_1}\right)\left(1+\frac{K_1}{k_2}\right)\left(1+\frac{K_2}{k_3}\right)\ldots\ldots\ldots\left(1+\frac{K_{p-1}}{k_p}\right)},$$

which reduces to equation (9) when $K_0$ and $K_p$ are each zero, which will always be the case if the initial and final spherical aberration be zero.

## §6. SPECIAL CASE WHEN THE REFRACTING SURFACE IS PLANE OR OF EXCEPTIONALLY LARGE RADIUS OF CURVATURE

When $r$ is very large or infinite, $c$ is also very large or infinite, while $\sin I$ is equal or nearly equal to $-\sin U$ and $\sin I'$ to $-\sin U'$. Hence $\Delta(UU')$ tends to vanish identically. The following substitute formula for $nuc\,\Delta(UU')$ may then be used:—

Since $nuc=nri$, and $\sin I=c\sin U/r$, $\sin I'=c'\sin U'/r$, we have, substituting in equation (5),

$$\begin{aligned} nuc(\sin U+\sin I-\sin I'-\sin U') &= nri\{\sin U(1+c/r)-\sin U'(1+c'/r)\}\\ &= ni(L\sin U-L'\sin U'), \qquad \ldots\ldots(10)\end{aligned}$$

the $r$s cancelling out.

This formula gives more accurate results for large radii and can be used for a plane surface. Otherwise it is less convenient as it involves two linear constants instead of one.

To conclude this section it may be observed that the advantages in the use of equation (6) will be found more apparent in the case of separated compound lenses than for thin systems of conventional type, where the possibilities of combination of glasses and surfaces have been thoroughly explored.

Even where trigonometrical methods are preferred, however, the quantity $nuc\Delta(UU')$ should still be formed for each surface parallel to the other work, since it is so readily calculated from the available elements. It shows at once how much each surface is contributing to the total spherical aberration, reveals the necessity for adjustments at an earlier stage, and shows to what extent the spherical aberration has accumulated beyond the power of subsequent surfaces to reverse.

## §7. CURVATURE OF FIELD FOR CENTRAL RAYS

The formation of images by thin pencils passing through the centre of curvature of each refracting surface can seldom be realized in practice. The Petzval surface on which such theoretical images would be formed is, however, of considerable practical importance, since the distance of the real intersections of the image-forming rays, before or behind this surface, is the measure of the degree of astigmatism introduced by the refracting system, and, as is well known, the formation of a plane anastigmatic image of a plane object requires as a first condition that the Petzval curvature shall also be zero.

The Petzval curvature for indefinitely thin pencils of central rays may be defined by the well-known formula

$$\frac{2x_q'}{n_q' Y_q'^2} = \frac{2x_1}{n_1 Y_1^2} + \sum_{p=q}^{p=1} \frac{n_p' - n_p}{n_p n_p' r_p},$$

where $r_p$ is the radius of curvature of the $p$th refracting surface, $x_1$ and $Y_1$ are the rectangular coordinates of an extra-axial object point, referred to an origin B on the optical axis where it is cut by the object surface, and $x_p'$ and $Y_p'$ are the corresponding rectangular coordinates of the $p$th image of the point $x_1 Y_1$ referred to the $p$th paraxial image of B as origin.

This equation is an approximation only, and although for most conjugate positions of object and image and a fairly small angle of field it gives sufficiently accurate results, with the development of wide-angle systems and of less conventional types there would appear to be good grounds for replacing, wherever possible, the approximate Seidel conditions with exact formulae. Coma is already taken care of by the sine condition; an exact addition equation for axial spherical aberration has been supplied in a previous section; summations are now given for the Petzval and sagittal curvatures.

An exact form of the Petzval equation for a single surface follows from equation (2*b*), viz.,

$$\Delta l' = \frac{n'}{n} m^2 \Delta l \left(\frac{1 - m_0 \Delta l}{f}\right).$$

For, following the well known method for finding the Petzval equation, with B as origin, let $x$, $Y$ (figure 2) be the coordinates of an extra-axial point in the

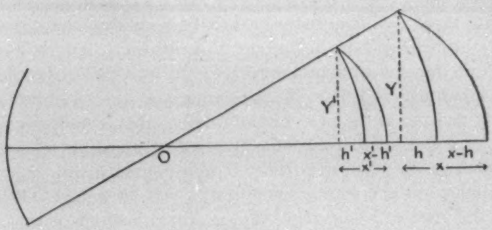


Figure 2.

object surface and $x'$, $Y'$ the coordinates of the corresponding extra-axial image point formed by an indefinitely thin pencil of central rays. With centre at the centre of curvature of the refracting surface, describe arcs through the points $xY$ and $x'Y'$. Let $h$ and $h'$ be the respective depths of curvature of these two arcs. We then have

$$\Delta l' = x' - h', \qquad \Delta l = x - h, \qquad Y'/Y = m.$$

Substituting in the above equation for $\Delta l'$ and $\Delta l$, we have

$$x' - h' = \frac{n'}{n} m^2 (x - h)\left(1 - m_0 \frac{x - h}{f}\right).$$

Dividing by $n'Y'^2 = n'm^2Y^2$ and transposing,

$$\frac{x'}{n'Y'^2} - \frac{x}{nY^2} = \frac{h}{Y}\left(\frac{1}{n'Y'} - \frac{1}{nY}\right) - m_0\frac{(x-h)^2}{nY^2f},$$

since

$$h/Y = h'/Y'.$$

Remembering that $Y'/Y = C'/C$ and $C'/Y' = C/Y$, and that $\dfrac{Y^2+h^2}{Y^2} = \dfrac{2Ch}{Y^2}$,

$$\frac{x'}{n'Y'^2} - \frac{x}{nY^2} = -\frac{1}{2}\frac{n'-n}{nn'r}\left(1 + \frac{h^2}{Y^2} - 2\frac{n'}{n}m_0\frac{(x-h)^2}{Y^2}\right).$$

Again, if $\theta$ is the semi-angle of field, $\dfrac{h}{y} = \tan\tfrac{1}{2}\theta$, so that, finally,

$$\frac{x'}{n'Y'^2} = \frac{x}{nY^2} - \frac{1}{2}\frac{n'-n}{nn'r}\left\{1 + \tan^2\tfrac{1}{2}\theta - 2\frac{n'}{n}m_0\left(\frac{x}{Y} - \tan\tfrac{1}{2}\theta\right)^2\right\}. \qquad \ldots\ldots(11)$$

Since the $x'/n'Y'^2$ of one surface corresponds to the $x/nY^2$ of the next, the exact position of any image point formed by indefinitely thin central pencils in the Petzval surface after $q$ refractions is given by the formula

$$\frac{x_q'}{n_q'Y_q'} = \frac{x_1}{n_1Y_1} - \frac{1}{2}\sum_{p=q}^{p=1}\frac{n_p'-n_p}{n_pn_p'r_p}\left\{1 + \tan^2\tfrac{1}{2}\theta_p - 2\frac{n_p'}{n_p}m_{0_p}\left(\frac{x_p}{Y_p} - \tan\tfrac{1}{2}\theta_p\right)^2\right\}. \qquad \ldots\ldots(12)$$

If the initial object surface is plane then $\dfrac{x_1}{n_1Y_1}$ is zero.

The right-hand side of equation (11) then becomes

$$\frac{1}{2}\frac{n'-n}{nn'r}\left\{1 + \tan^2\tfrac{1}{2}\theta\left(1 - 2\frac{n'}{n}m_0\right)\right\}.$$

This can vanish:

(*a*) For a plane refracting surface, i.e. $r = \infty$.

(*b*) When the object is at the centre of curvature of the refracting surface, i.e. when $m = n/n'$ and $\tan\tfrac{1}{2}\theta = 1$.

Combining these two conditions, it will be seen that the virtual image of a plane object formed by a plano-convex lens of focal length $F$ at a distance $(n'-n)F/n'$ from the object has no Petzval curvature, although by the approximate formula the curvature is $1/nF$. This relative position of object and image is approximately realized at the field lens of a Huygenian eyepiece and in the front combination of some microscope objectives, thus assisting in the flattening of the field in such cases. Unfortunately, in a cemented plano-convex achromatic lens, the tendency of the crown is to neutralize this advantage, unless the refractive index of the glass of lower dispersion is at least as high as that of the flint.

Equation (11) consists of a principal term representing the usual Petzval sum with the addition of a corrective term for each surface. However, the exact condition that the Petzval sum may be zero can be expressed in a simpler form by dividing the axial magnification formula (1) by $Y'$ instead of by $Y'^2$ and replacing $\Delta l$ and $\Delta l'$ by $(x-h)$ and $(x'-h')$ as before.

Then for a single surface we have

$$u'\frac{x'}{Y'} = u\frac{x}{Y} + (u'-u)\frac{h}{Y} = u\frac{x}{Y} - \delta\tan\tfrac{1}{2}\theta, \qquad \ldots\ldots(13)$$

where $\delta$ is the paraxial deviation for a ray of semi-aperture $u$ and $u'$ before and after refraction.

For a succession of spherical surfaces the formula becomes

$$u_q' \frac{x_q'}{Y_q'} = u_1 \frac{x_1}{Y_1} - \sum_{p=q}^{p=1} \delta_p \tan \tfrac{1}{2}\theta_p, \qquad \ldots\ldots(14)$$

so that if the object and final image surfaces are each to be plane,

$$\sum_{p=q}^{p=1} \delta_p \tan \tfrac{1}{2}\theta_p = 0. \qquad \ldots\ldots(14a)$$

## §8. ADDITION FORMULA FOR SAGITTAL ASTIGMATISM

By using equation (3) a condition similar to equation (13) above may be obtained for the curvature of the sagittal field.

Dividing by $n'Y' \sin V' = nY \sin V$, and again substituting $(X'-H')$ for $\Delta L'$ and $(X-H)$ for $\Delta L$, and transposing, we have

$$\begin{aligned} u'\frac{X'}{Y'} &= u\frac{X}{Y} + (u'-u)\frac{H}{Y} + \frac{n'u'c'\,\Delta(VV')}{nY \sin V} \\ &= u\frac{X}{Y} - \delta \tan \tfrac{1}{2}\Psi + \frac{n'u'c'\,\Delta(VV')}{nY \sin V}, \end{aligned} \qquad \ldots\ldots(15)$$

using capital $X$ and capital $H$, $V$, $V'$ and $\Psi$ in place of $x$, $h$, $u$, $u'$ and $\theta$ to distinguish the sagittal ray.

Then for a succession of $q$ surfaces

$$u_q' \frac{X_q'}{Y_q'} = u_1 \frac{X_1}{Y_1} - \sum_{p=q}^{p=1} \delta \tan . \tfrac{1}{2}\Psi + \sum_{p=q}^{p=1} \frac{n_p'u_p'c_p'\,\Delta(VV')_p}{n_p Y_p \sin V_p}. \qquad \ldots\ldots(16)$$

The final term in equation (16) has one serious disadvantage. It has already been stated that in the spherical-aberration formula (5), the scale of each element of the complete system is determined by a single linear constant $c'$. This constant again appears in equation (16), but there is no longer complete freedom of selection.

Although the scale of the system as a whole is still open, the relative scale as between the separate components has already been pre-determined by the selection (in conjunction with the paraxial deviation $\delta$) of the angular subtense $\Psi_p$ of the field presented to each surface, thus fixing the positions of the centres of curvature and leaving no remaining latitude in the radii after these have produced the prescribed deviations.

It is therefore desirable to expand this term as follows:—

$$\begin{aligned} \frac{n'u'c'\,\Delta(VV')}{n'Y' \sin V'} &= \frac{nuc\,\Delta(VV')}{nY \sin V} = u\frac{\Delta(VV')}{\sin V}\left(\frac{C+X-H}{Y}\right) \\ &= u\frac{\Delta(VV')}{\sin V}\left(\operatorname{cosec} . \Psi + \frac{X}{Y} - \tan \tfrac{1}{2}\Psi\right). \end{aligned} \qquad \ldots\ldots(17)$$

For the preliminary work it will be sufficient to take $\dfrac{u}{\sin V}\Delta(VV') \operatorname{cosec} \Psi$ as the approximate value of the above term. The remaining corrections can be applied at a later stage.

Any combination of the second and third terms on the right-hand side of equation (16) will give a flat field for sagittal rays provided the total sum is zero; but to ensure that the sagittal image is formed in the Petzval surface, equation (14) must also be satisfied; or, a little less precisely, the two sums in equation (16) must be separately zero.

# DYNAMIC MEASUREMENT OF YOUNG'S MODULUS FOR SHORT WIRES

By Y. L. YOUSEF,
Fouad I University, Cairo

*MS. received* 3 *April* 1945

*ABSTRACT.* The method consists essentially in the measurement of the time for a quarter vibration of a short wire clamped at one end. The wire breaks a circuit as it starts its transverse vibration, and is made to strike a very light lever and open another circuit just as it completes one-quarter of its first vibration. The time between the two events is measured ballistically. From a knowledge of the time of vibration, the length, the diameter and the density of the wire, Young's modulus can be readily calculated. The method is particularly suitable for thin wires having a diameter of the order of 0·015 cm. and a length of about 2 cm. The results are consistent to a few per cent. Comparison of the results with the accepted values for five different wires shows a good agreement.

---

## §1. INTRODUCTION

LORD RAYLEIGH (1926) correlated the frequency of transverse vibrations of a bar with its Young's modulus, and the correlation was used, probably for the first time, by Prosad (1929), who maintained the vibrations electrically, and determined the frequency chronographically by connecting to the free end an attachment for which an end-correction was applied. This correction amounted to more than 200% for lengths of the order of several cm. The mean results of Prosad differ by about 5% from the results he obtained by the flexural method, while his curves suggest that a divergence of about 15% from the average is possible.

In the present work, the frequency is calculated from the time of one-quarter of a vibration of the wire.

## §2. EXPERIMENTAL DETAILS

A length of a few cm. of a thin straight wire is used. One end of the wire is rigidly clamped, and the other end can execute lateral vibrations. It is arranged that as the free end of the wire is released from a depressed position it breaks a circuit, and just as it reaches its normal position it opens another circuit. The time interval between the two events is measured by a condenser circuit described by Klopsteg (1920).

The principle will be understood by reference to figure 1. The two contacts $K_1$ and $K_2$ are opened in quick succession at the beginning and the end of the short interval to be measured. Then the battery is reversed, $K_1$ short-circuited by $K_3$, and the two-way key $K_5$ put to the galvanometer side. If the setting of the resistances $R$ and $S$ has been right, the galvanometer throw will be zero, and the interval will be given by

$$t = CR \log_e R/S.$$

*The contact $K_1$.* The wire X under test is clamped at C (figure 2) by means of two plane-edged aluminium jaws which can be tightened on the wire by nut and screw. The rigid clamp forms one terminal A of the contact. The clamped end of the wire is sharply defined by the plane edge of the clamp. The free end K of the wire can be depressed and kept in a constrained position by means of a stout nickel wire D having a small, flat, horizontal edge, with which it makes good electrical contact. D is connected to the second terminal B.

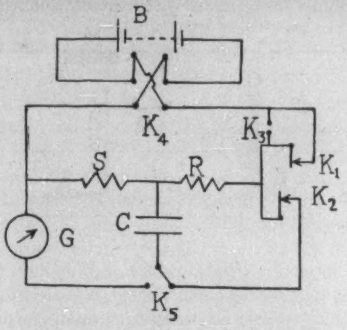


Figure 1. The principle of the method.

The contact at K is made at the extreme edges of X and D, so that when D is suddenly pulled outwards by means of an electromagnet or an attached thread, the break of the battery-circuit of figure 1 is abrupt and the current drops immediately to zero.

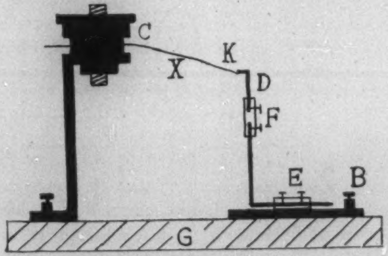


Figure 2. The contact $K_1$.

The position of D can be adjusted both horizontally and vertically by means of the screws E and F, and thus measurements can be carried out for various lengths of X and for various amplitudes of vibration.

The arrangement is fitted on an ebonite base G, which itself is fixed centrally on a levelling table.

*The contact $K_2$.* The contact $K_2$ has a very small inertia and at the same time has a reasonably low and fairly constant resistance (0·7 to 0·9 ohms).

JOL (figure 3) is a fine steel wire (diam. = 0·02 cm., length = 5 cm.) pivoted at O, where a pin passes through a small loop in the wire. The pin is soldered to a copper plate P connected to a terminal screw $T_1$. The end L is bent at right angles to the length JL, and a short thin platinum wire (diam. = 0·0075 cm.) is soldered at N, where the lever rests on the plane top of the copper wire S, which is freshly but thinly amalgamated with mercury, and which forms the second lead. A very thin, loose wire W of copper (S.W.G. 40) is soldered between JL and the plate P to improve the contact between the lever and the terminal $T_1$. It also helps, by the slight pressure it exerts, in keeping N and S in good contact, and thus the resistance between $T_1$ and $T_2$ is small. The whole arrangement is mounted on an ebonite base E, which is held by a fixed stand.

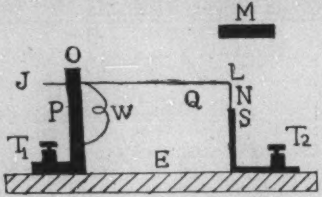


Figure 3. The contact $K_2$.

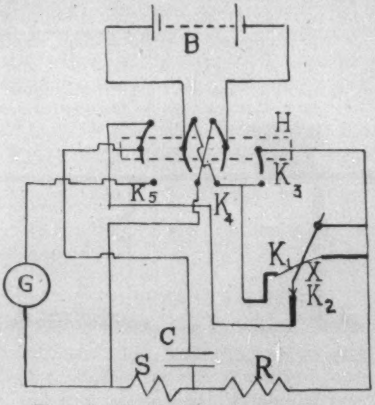


Figure 4. The experimental arrangement.

The vibrating wire X (figure 2) passes between E and JL, and is arranged to strike the lever normally at some point Q near the end L. As the lever receives the impulse, the contact between N and S is suddenly broken, and the point L is raised enough to come into the field of action of the small magnet M, to which it is attracted, and thus the contact remains open.

A force of a few centigrammes weight is sufficient to pull off the point N from S.

*The actual arrangement.* For $K_3$, $K_4$ and $K_5$ of figure 1 we used mercury keys of the rocking type. They are constructed on one block of paraffin wax, and are all operated by a common handle H (figure 4). The lengths of the movable links between the mercury pools are such that on switching over from the first position to the second, the short-circuiting key $K_3$ and the reversing key $K_4$ are closed just before the discharge $K_5$.

R and S are two large resistance boxes ; C is a standard mica condenser of capacity $2{\cdot}0005 \times 10^{-6}$ F. ; G is a ballistic galvanometer having a sensitivity of 2 scale divisions per micro-coulomb, and a deflection of about 1/10 of a scale division can be detected ; the battery B has a constant e.m.f. of about 32 volts.

*Adjustment of the test wire.* The position of the unconstrained test wire, which is situated normally with its free end below the end of the lever, is carefully adjusted by levelling until it starts to raise the lever, and the contact $K_2$ is just broken. The break may be judged by viewing the tip N and its image on the bright amalgamated surface of S through a lens (figure 3). A white background assists in producing a distinct view. The right position may be located to within much less than corresponds to a 20° rotation by one of the screws of the levelling table supporting $K_1$. This rotation is equivalent to a linear displacement of less than 1/200 of the smallest amplitude of vibration we used, which is about 3 mm.

On account of the very small inertia of the lever, the test wire is not appreciably bent as it starts to raise the lever. This is especially so for short lengths of the test wire, since the bending varies directly as the cube of the length.

When the "zero" position of the test wire is exactly located, it is depressed to the point D (figure 2) to close the contact $K_1$. Care is always taken that the depression of the free end does not exceed 1/6 of the length used. For this depression, the initial amplitude of vibration may be regarded as small, and the wire will not change its zero position.

By virtue of the large velocity of vibration near the mean position, an error of $z$ % of the amplitude arising from a slight error in locating the test wire will produce one of only $\frac{2}{\pi}z$ % in the time of one-quarter vibration, or an error of $\frac{4}{\pi}z$ % in Young's modulus.

## § 3. THE METHOD OF MEASUREMENT

With $K_2$ closed (figure 4), observations are made by first approximately adjusting the position of the test wire, and then closing $K_1$ and taking trial values for $R$ and $S$. On releasing the test wire, the contacts $K_1$ and $K_2$ will be opened in succession after a quarter vibration. The keys $K_3$, $K_4$ and $K_5$ respectively are then immediately switched over by the handle H. If G shows a deflection, readjustment of the resistances $R$ and $S$ is made until the deflection is reduced practically to zero. It is generally sufficient to change one of the resistances only, and keep the other constant at a suitable value. The test wire is then exactly located, and the final adjustment of the resistance is made to give zero deflection exactly.

The values of $R$ and $S$ are now recorded, and a new balance is obtained with other values of the resistances, the zero position of the test wire being always checked during the final adjustments. An average time can thus be obtained.

If the contact $K_2$ has a small resistance $r$, then

$$t = C(R+r)\log_e \frac{R}{S}.$$

The frequency of transverse vibrations will be given by

$$\nu=\frac{1}{4t}.$$

Lord Rayleigh (loc. cit.) has shown that, for lateral vibrations of a bar of length $l$, density $\rho$ and Young's modulus $Y$, the frequency is given by

$$\nu=\frac{km^2}{2\pi l^2}\sqrt{(Y/\rho)},$$

where $k$ is the radius of gyration of the section about an axis perpendicular to the plane of bending, and $m$ is an abstract number whose value for the gravest mode of vibration of a clamped-free rod is 1·8751.

For a circular wire of radius $a$, $k=\frac{1}{2}a$, and hence

$$Y=\frac{12{\cdot}77\rho}{a^2}\nu^2l^4,$$

from which it will be seen that for a given wire $\nu^2l^4$ is a constant quantity.

The length of the wire between the clamped and the free end is measured by callipers, and its average diameter is measured by a micrometer screw gauge. The density is determined from the mass and the dimensions of the sample; frequency determinations are made for various lengths of the wire, and the mean value of $\nu^2l^4$ is obtained and used for calculating $Y$.

## §4. RESULTS

Typical results are given in tables 1, 2 and 3, where the c.g.s. units are used throughout. The data in table 1 show the accuracy with which the

Table 1. Accuracy in measuring the time

| $R$ (ohms) | $S$ (ohms) | $t$ (sec.) | Average $t$ | % deviation |
|---|---|---|---|---|
| 1500 | 699 | 0·00229 | 0·00226 | −1·4 |
| 1250 | 514 | 0·00222 | | +1·8 |
| 1000 | 322 | 0·00227 | | −0·4 |
| 750 | 168 | 0·00225 | | +0·4 |

Table 2. Effect of change of amplitude

| $A/l$ | $R$ (ohms) | $S$ (ohms) | $t$ (sec.) | Mean $t$ |
|---|---|---|---|---|
| 0·222 | 2000 | 1010 | 0·002734 | 0·002738 |
| | 1500 | 602 | 0·002741 | |
| 0·109 | 2000 | 1013 | 0·002723 | 0·002715 |
| | 1500 | 609 | 0·002707 | |

Table 3. Results of the measurements

| Material | $\rho$ (gm./c.c.) | $a$ (cm.) | $l$ (cm.) | $R$ (ohms) | $S$ (ohms) | $\nu$ (sec.$^{-1}$) | $\nu^2 l^4 . 10^{-5}$ (c.g.s. units) | $Y_0 . 10^{-11}$ (c.g.s. units) | $Y_a . 10^{-11}$ (c.g.s. units) |
|---|---|---|---|---|---|---|---|---|---|
| Brass | 8·604 | 0·02283 | 5·29 | 2000 | 914 | 79·74 | 49·8 | 10·3 | 9·7 to 10·2 |
| | | | 4·52 | 1500 | 699 | 109·0 | 49·6 | | |
| | | | 3·33 | 1000 | 535 | 199·5 | 48·9 | | |
| | | | 2·68 | 500 | 219 | 302·0 | 47·1 | | |
| Steel | 7·997 | 0·02793 | 5·32 | 1000 | 414 | 141·4 | 160·2 | 20·7 | 19·5 to 20·6 |
| | | | 4·68 | 1000 | 500 | 180·1 | 155·5 | | |
| | | | 3·63 | 500 | 218 | 300·6 | 156·9 | | |
| | | | 2·69 | 500 | 318 | 551·3 | 159·5 | | |
| Constantan | 8·845 | 0·0273 | 5·535 | 1000 | 322 | 110·2 | 114·0 | 16·7 | 16·3 |
| | | | 5·385 | 1000 | 335 | 114·2 | 109·0 | | |
| | | | 4·06 | 750 | 333 | 204·9 | 114·1 | | |
| | | | 3·60 | 500 | 186 | 252·2 | 106·9 | | |
| Nickel | 8·919 | 0·008675 | 2·605 | 750 | 273 | 164·8 | 12·50 | 19·6 | 20·2 |
| | | | 2·47 | 750 | 313 | 190·4 | 13·49 | | |
| | | | 2·19 | 750 | 372 | 237·7 | 12·99 | | |
| | | | 1·945 | 600 | 299 | 298·6 | 12·76 | | |
| Tungsten | 19·8 | 0·00725 | 3·29 | 1500 | 564 | 85·9 | 8·48 | 39·9 | 39·0 |
| | | | 2·93 | 1000 | 306 | 105·5 | 8·19 | | |
| | | | 2·74 | 1000 | 355 | 120·6 | 8·20 | | |
| | | | 2·19 | 1000 | 519 | 190·4 | 8·32 | | |
| Nichrome | 8·399 | 0·01802 | 5·13 | 1500 | 609 | 92·3 | 58·9 | 19·7 | ? |
| | | | 3·41 | 1000 | 555 | 211·7 | 60·6 | | |
| | | | 2·84 | 500 | 218 | 305·6 | 60·7 | | |
| | | | 2·62 | 500 | 246 | 351·6 | 58·2 | | |

adjustments can be made and the time measured. The observations were made on a constantan wire of length 5·535 cm. and radius 0·0273 cm.

The maximum departure from the average is less than 2%.

The figures in table 2 show the effect of change of amplitude on the time of a quarter vibration. The observations were made on a nichrome wire of length 5·13 cm. and radius 0·01812 cm. The times for two widely different ratios of amplitude ($A$) to length ($l$) are given.

It will be seen that for the ratios of $A/l$ given, the effect of change of amplitude does not exceed 1 %, which is of the same order of magnitude as the experimental error. However, the ratios of $A/l$ we actually used in our determinations were always less than 0·16.

A summary of typical results at room temperature (18° C.) for six different wires is given in table 3, where $Y_0$ stands for Young's modulus by the author's method, and $Y_a$ stands for the accepted value (Kaye and Laby, 1928; Childs, 1943).

The table shows how far the results of different observations on the same wire are concordant among themselves. It will be also noticed that, in general there is a good agreement between the observed and the accepted values of the elasticity.

## § 5. CONCLUSION

When measures are taken to secure good clamping of the test wire by a massive support, small inertia of the lever, high electrical conductivity at the contacts $K_1$ and $K_2$, small area of contact with the tip of the lever, exact location of the test wire, good insulation of the various keys, and sudden break of the battery circuit by $K_1$, the method appears to be very useful in measuring, with ordinary laboratory equipment, Young's modulus of elasticity, particularly for thin and short wires. No correction is required, and the wire need not be especially shaped or adapted. The average result will give a good measure of the true modulus of elasticity.

I wish to express my thanks to Dr. M. A. El-Sherbini for his valuable help and advice during the progress of this work.

## REFERENCES

Childs, W., 1943. *Physical Constants* (Methuen), p. 18.
Kaye, G. and Laby, T., 1928. *Tables of Physical and Chemical Constants*, p. 29.
Klopsteg, P., 1920. *Phys. Rev.* **15**, 12.
Prosad, K., 1929. *Phil. Mag.* **7**, 548.
Rayleigh, Lord, 1926. *Theory of Sound*, **1**, 273.

# A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: III. THE CONTRIBUTIONS TO THE IMAGE ABERRATIONS MADE BY THE INDIVIDUAL SURFACES OF A LENS SYSTEM

By F. D. CRUICKSHANK,
University of Tasmania

*MS. received* 26 *February* 1945

*ABSTRACT*. A method is described which permits the computation of the contributions made by the individual surfaces of a lens system to the tangential aberrations of the final image for a pencil of finite aperture and obliquity. It employs the transfer coefficients previously described and forms part of a general development of the trigonometrical method of design. When applied to pencils of small aperture and small obliquity it gives an analysis of the surface contributions in exact agreement with known primary methods, and thus provides an extension from the analysis of the primary to the trigonometrically determined aberrations. In particular, expressions are derived for the contributions to the chromatic aberrations made by any component of the system which are simple functions of the dispersion of the component.

---

## §1. INTRODUCTION

In the course of the design of an optical system it is frequently of considerable value to know the contributions which each surface makes to the image aberrations. Methods of computing the surface contributions to the primary aberrations are well known, but, for aberrations other than these, no general analysis into surface contributions is available. Using the properties of the differential transfer coefficients developed in Part I of this paper, it is possible to solve this problem for the tangential aberrations of a system. We proceed to develop a trigonometrical method of computing the surface contributions characteristic of a lens system for a given aperture and position in the image field.

## §2. THE CONTRIBUTIONS TO THE LONGITUDINAL AND TRANSVERSE CHROMATIC ABERRATIONS

In Part I an expression is derived for the longitudinal chromatic aberration of the system at a given zone in the form

$$Lch' = \Sigma \frac{\partial L_d'}{\partial N_h}(N_r - N_v)_h, \qquad \text{......(67)}$$

where

$$\frac{\partial L_d'}{\partial N_h} = \frac{\partial L_d'}{\partial n_2}\frac{1}{N_{h+1}} - \frac{\partial L_d'}{\partial n_1}\frac{N_{h-1}}{N_h^2}, \qquad \text{......(68)}$$

the summation being extended over all the components of the system. Each term in the summation in (67) represents the contribution made by a component to the longitudinal chromatic aberration of the final image. Thus for the general component, *h*, the trigonometrically determined contribution to the longitudinal chromatic aberration, which we denote by $TLchC_h{}'$, is given by

$$TLchC_h{}' = \frac{\partial L_d{}'}{\partial N_h}(N_r - N_v)_h. \qquad \ldots\ldots(69)$$

If the surface contributions are required rather than the component contributions, it is obvious that they may be easily obtained. Each term in the summation of equation (69) is derived from a pair of terms as seen in equation (68), the two members of each pair referring to single surfaces. Hence the longitudinal chromatic aberration may be expressed as a summation over the individual surfaces of the system if desired. For each air-glass surface there is a single term, while for a cemented surface there are two terms each of which involves the partial dispersion of one of the glasses forming the contact surface. The term, or the sum of the two terms, relating to each surface gives the contribution of that surface to the final value of $Lch'$. For the paraxial chromatic aberration we have similarly

$$lch' = \Sigma \frac{\partial l_d{}'}{\partial N_h}(N_r - N_v)_h, \qquad \ldots\ldots(70)$$

whence

$$TlchC_h{}' = \frac{\partial l_d{}'}{\partial N_h}(N_r - N_v)_h. \qquad \ldots\ldots(71)$$

In Part I an expression was also derived for the transverse chromatic aberration for a given obliquity in the form

$$Tch' = \Sigma \left[\frac{\partial H'_{prd}}{\partial N_h}\right]_{L'}(N_r - N_v)_h. \qquad \ldots\ldots(72)$$

This expression is based on the trace of a principal ray of the desired obliquity using the refractive indices of the intermediate colour *d*. The subscript $L'$ indicates that the plane in which the ordinates, $H'$, are measured is a fixed plane. It is convenient to use the paraxial image plane for this purpose. In a manner similar to that followed in the case of the longitudinal aberration, it may be shown that the trigonometrically determined contribution to the transverse chromatic aberration for the general component, *h*, is given by

$$TTchC_h{}' = \left[\frac{\partial H'_{prd}}{\partial N_h}\right]_{L'}(N_r - N_v)_h. \qquad \ldots\ldots(73)$$

If the surface contributions are desired, the equation (72) may be rearranged as a summation over the surfaces in a manner similar to that already discussed in the case of the longitudinal aberration.

The contributions made by the components of the system to the final chromatic aberrations are frequently of great value in practical designing, being much more useful than the contributions made by the surface of the system. Their value lies in the fact that they are such simple functions of the dispersions, and it is often possible to improve the achromatism of the system by a change of glass at one or more components after inspection of the values of the contributions.

In the accompanying table, the results of the computation of the contributions made by the surfaces of a photographic objective to the paraxial longitudinal chromatic aberration and to the transverse chromatic aberration at two obliquities are compared with the corresponding values for the contributions to the *primary* chromatic aberrations as calculated by the method of Conrady (1929).

Table

| Surface | *TlchC'* | Conrady *lchC'* | *TTchC'* | Conrady *TchC'* | *TTchC'* | Conrady *TchC'* |
|---|---|---|---|---|---|---|
| | | | 4° | | 10° | |
| 1 | 3·5406 | 3·5406 | −0·0086 | −0·0083 | −0·0223 | −0·0210 |
| 2 | 3·8759 | 3·8759 | 0·0778 | 0·0767 | 0·2138 | 0·1931 |
| 3 | −6·6646 | −6·6646 | −0·1331 | −0·1313 | −0·3659 | −0·3304 |
| 4 | 0·4251 | 0·4251 | 0·0723 | 0·0713 | 0·1949 | 0·1795 |
| 5 | −2·0300 | −2·0300 | 0·0101 | 0·0099 | 0·0271 | 0·0250 |
| 6 | 0·0052 | 0·0052 | −0·0124 | −0·0122 | −0·0341 | −0·0308 |
| 7 | 1·6591 | 1·6591 | 0·0034 | 0·0034 | 0·0089 | 0·0086 |

In the first place it will be seen that the contributions determined by the two methods lead to identical results for the longitudinal aberration. Secondly, the table shows that at an obliquity of 4° the values given by the two methods are in fairly close agreement, the small difference being due to aberrations of higher order than the primary, which are already present. As the obliquity increases, the two sets of values diverge further, as is to be expected. At very small obliquities the two sets of values would be in exact agreement. It will thus be seen that a perfectly general trigonometrical method of calculating the surface and component contributions for the chromatic aberrations of a system has been established. It is accurate, as may be checked by full ray-traces, and when applied to the paraxial region and to pencils of small obliquity it gives results in exact agreement with existing primary theory, thereby becoming a satisfactory extension of known methods.

## § 3. THE SURFACE CONTRIBUTIONS TO THE SPHERICAL ABERRATION

In the ray-trace of the axial pencil we have traced rays incident at certain zones of the lens system. In the image formed after refraction at surface $i$ of the system, there is, for any such zone, a spherical aberration given by

$$\mathrm{LA}_i' = l_i' - L_i',$$

which may be expressed in angular measure as

$$\begin{aligned}\mathrm{AA}_i' &= \mathrm{LA}_i' \sin U_i'/S_i' \\ &= \mathrm{LA}_i' \sin U_i' \cos U_i'/(L_i' - X_i'). \qquad \ldots\ldots(74)\end{aligned}$$

The angle $\mathrm{AA}_i'$ measures the departure of the ray from the ideal direction after refraction at the surface $i$. If the ray could be turned through an angle $dU_i' = -\mathrm{AA}_i'$, the spherical aberration behind the surface $i$ would be reduced to zero, and the effect at the final image would be the introduction of an amount of spherical aberration opposite to that introduced by the actual refractions over the first $i$ surfaces of the system. In other words, the shift of the intersection point of the axial ray with the principal axis in the final image space due to this imagined rotation of the ray through the angle $dU_i' = -\mathrm{AA}_i'$, at surface $i$, will provide a measure of the sum of the spherical contributions of the first $i$ surfaces of the system. We shall denote this quantity by the symbol $\overset{i}{\Sigma}\, TSC'$, the letters standing for the trigonometrically determined spherical contribution. Hence

$$\overset{i}{\Sigma}\, TSC' = \frac{\partial L_k'}{\partial U_i'}\, dU_i' = -\frac{\partial L_k'}{\partial U_i'}\, \mathrm{AA}_i', \qquad \ldots\ldots(75)$$

and the individual contributions from the surfaces are given by

$$TSC_i' = \overset{i}{\Sigma}\, TSC' - \overset{i-1}{\Sigma}\, TSC'. \qquad \ldots\ldots(76)$$

This gives a general trigonometrical method of analysing the surface contributions to the final spherical aberration. Little additional calculation is involved when the general methods outlined in these papers are employed, for the necessary transfer coefficients are already obtained in the main computation. The only limitation to the accuracy of the method is that differential transfer coefficients are used to calculate the effects in the final image-space of the rotations, $dU'$, of the ray. If the latter are large, the transfer coefficients will not predict their effects accurately. This difficulty is met to a very large extent by using appropriate second-order correction terms, as developed in the previous paper. If $dU_k'$ is the change in the direction of the final emergent-ray due to the rotation of the ray through $dU_i'$ behind surface $i$, then the shift of the intersection point along the principal axis is by equation (46):

$$\delta L_k' = C(U_i') \operatorname{cosec} U_k'\, dU_i'(1 - \cot U_k'\, dU_k'), \qquad \ldots\ldots(77)$$

and with due attention to the sign of the longitudinal aberration this leads to

$$\overset{i}{\Sigma}\, TSC' = C(U_i') \operatorname{cosec} U_k'\, dU_i'(1 - \cot U_k'\, dU_k'). \qquad \ldots\ldots(78)$$

This equation (78), incorporating the correction term, should take precedence over the earlier equation (75) and be used for the normal calculation of $\overset{i}{\Sigma}\, TSC'$. The separate surface contributions then follow from equation (76).

## §4. THE SURFACE CONTRIBUTIONS TO THE DISTORTION

From the results of the trace of an oblique pencil through the lens system we can obtain the value of the distortion for this obliquity in the images formed after refraction at the successive surfaces of the system. In the usual linear measure this is

$$(\mathrm{dist}')_i = (H_{id}' - H_{prf}')_i,$$

and in angular measure

$$\begin{aligned}\mathrm{AD}_i' &= \mathrm{dist}_i' \cos U_{pri}' / S_{prfi}' \\ &= \mathrm{dist}_i' \cos^2 U_{pri}' / (l_i' - X_{pri}'). \qquad \ldots\ldots(79)\end{aligned}$$

This angle, $\mathrm{AD}_i'$ measures the departure of the principal ray from the ideal direction, as regards distortion, after refraction at surface $i$. If we could swing the principal ray through an angle $dU_{pri}' = -\mathrm{AD}_i'$, we should restore the ray to its ideal direction and thereby remove from the final image the amount of distortion introduced by the refractions at the first $i$ surfaces of the system. The effect of such an imagined rotation then is to introduce into the final image an amount of distortion opposite to that contributed by the first $i$ surfaces of the system, an amount which we denote by $-\overset{i}{\Sigma}\, TDC'$. Hence we have

$$-\overset{i}{\Sigma}\, TDC' = d(H'_{id} - H'_{prf}) = -dH'_{prf}$$

$$= -\frac{\partial H'_{prf}}{\partial U'_i}\, dU'_{pri} = \frac{\partial H'_{prf}}{dU'_i}\, \mathrm{AD}'_i,$$

that is,

$$\overset{i}{\Sigma}\, TDC' = -\frac{\partial H'_{prf}}{\partial U'_i}\, \mathrm{AD}'_i, \qquad \ldots\ldots(80)$$

and the contribution from the surface $i$ is given by

$$TDC'_i = \overset{i}{\Sigma}\, TDC' - \overset{i-1}{\Sigma}\, TDC'. \qquad \ldots\ldots(81)$$

Equation (80) is adequate provided the angles $dU'_{pri}$ are small. Improved accuracy is obtained by using the second-order correction term ; the difference in the amount of computation involved is so little that the corrected equation (82) given below should be used as a matter of course. Using equation (49) of the previous paper, and omitting the subscript *pr*, as all the symbols refer to the principal ray, we obtain finally

$$\overset{i}{\Sigma}\, TDC' = C(U_i')_{prf} \sec U_k'\, dU_i'(1 + \tan U_k'\, dU_k'). \qquad \ldots\ldots(82)$$

## § 5. THE SURFACE CONTRIBUTIONS TO THE TANGENTIAL COMA

We begin by computing the value of the tangential coma for a pencil of given obliquity in the images formed after refraction at the successive surfaces of the system. If we choose to regard the $a$ and $b$ rays of the pencil as aberrant, then it is the displacement of their intersection point, Ab, from the intersection point Pr which afflicts the system with coma. If after refraction at surface $i$ we could swing the $a$ and $b$ rays until both passed through the point Pr, the tangential coma originally present in the image behind surface $i$ would be removed. Such a swing of the rays would also leave the curvature of the tangential field unchanged. The effect of such an imagined rotation of the rays on the final image would be the introduction of an amount of tangential coma opposite to that contributed by the first $i$ surfaces of the system, that is, an amount which we denote by $-\overset{i}{\Sigma}\, TCC'$. Thus we have a basis for the derivation of equations for computing the contribution of each surface of the system to the tangential coma of the final image.

The angle through which the $a$ ray must be rotated is

$$dU'_{ai} = -\mathrm{coma}'_{Ti} \cos U'_{ai}/S'_{ai}$$

$$= -\mathrm{coma}'_{Ti} \cos^2 U'_{ai}/(L'_{ab} - X_a')_i. \qquad \ldots\ldots(83)$$

The corresponding angle for the rotation of the $b$ ray is

$$dU'_{bi} = -\mathrm{coma}'_{Ti} \cos^2 U'_{bi}/(L'_{ab} - X_b')_i. \qquad \ldots\ldots(84)$$

In order to calculate the contributions at the final image we must follow the shifts of the intersection points Ab and Pr in the final image space due to these rotations. The shifts are

$$dL'_{ab} = \frac{\partial L'_{ab}}{\partial U'_{ai}}\, dU'_{ai} + \frac{\partial L'_{ab}}{\partial U'_{bi}}\, dU'_{bi}, \qquad \ldots\ldots(85)$$

$$dH'_{ab} = \frac{\partial H'_{ab}}{\partial U'_{ai}} dU'_{ai} + \frac{\partial H'_{ab}}{\partial U'_{bi}} dU'_{bi}, \qquad \ldots\ldots(86)$$

$$dH'_{pr} = -dL'_{ab} \tan U'_{prk}. \qquad \ldots\ldots(87)$$

It then follows that

$$\sum^{i} TCC' = dH'_{ab} - dH'_{pr}, \qquad \ldots\ldots(88)$$

and the contribution made by the surface *i* is given by

$$TCC'_i = \sum^{i} TCC' - \sum^{i-1} TCC'. \qquad \ldots\ldots(89)$$

As in the previous case, it is generally advisable to use the second-order correction terms. It will be easily seen that the modifications necessary in equations (85) to (87) are that the first term on the right-hand side of (85) and (86) is multiplied by $[1 - \cot(U_a' - U_b')_k dU'_{ak}]$, and the second term by the corresponding quantity $[1 - \cot(U_a' - U_b')_k dU'_{bk}]$.

The essence of the foregoing method is the computation of the effect at the final image of a change made in the refracted rays behind surface *i*, the change being such that the coma present behind this surface is removed. Another change which would remove the coma in the image behind the surface *i* is a swing of the principal ray until it passes through the point $\mathrm{Ab}_i$. It may seem more logical to make the changes in the directions of the rays *a* and *b*, because we regard coma as due to the aberrant behaviour of the outer rays of the pencil, but it is simpler to make the change of the principal ray for the purpose of calculating the coma contributions, as it involves less than half the amount of computation. The fact that the contributions computed either way are in close agreement confirms the validity of the general method. The angle through which the principal ray must be rotated to remove the coma behind surface *i* is

$$\begin{aligned} dU'_{pri} &= \mathrm{coma}'_{Ti} \cos U'_{pri} / S'_{pri} \\ &= \mathrm{coma}'_{Ti} \cos^2 U'_{pri} / (L'_{ab} - X'_{pr})_i. \qquad \ldots\ldots(90) \end{aligned}$$

Following out the effect in the final image space of such a rotation we have

$$\sum^{i} TCC' = -\frac{\partial H'_{pr}}{dU'_{pri}} \partial U'_{pri}. \qquad \ldots\ldots(91)$$

An additional advantage of the second method is the ease with which the correction term may be applied. As in the case of distortion, we obtain

$$\sum^{i} TCC' = -C(U_i')_{pr} \sec U'_{prk} dU'_{pri} (1 + \tan U_k' dU'_{prk}). \qquad \ldots\ldots(92)$$

Finally, then, we use equations (90), (92) and (89) for the computation of the contributions to the tangential coma.

## §6. THE SURFACE CONTRIBUTIONS TO THE CURVATURE OF THE TANGENTIAL FIELD

The curvature of the tangential field is measured usually by the distance of the intersection point Ab from a plane at right angles to the axis through the focus of some chosen axial ray. This ray is usually the extreme ray of an axial pencil having the same relative aperture as the oblique pencil. Thus

$$X_T' = L'_{ab} - L_m'.$$

As long as the point Ab lies in the plane through the intersection point of the ray *m* with the principal axis, the curvature of the tangential field would be zero according to the last equation, but the coma would be profoundly affected by the $H'$ co-ordinate of the point. Considering the curvature as quite distinct from the coma, we could say that in a comatic system the ideal location of the point Ab as far as curvature is concerned is in the plane of the *m* focus at a point distant $(\text{coma})_T'$ from the intersection point, *Prm*, of the principal ray with this plane. This analysis of coma and curvature may appear somewhat arbitrary, but so are most measures and analyses of the aberrations of a system. We propose to use the above ideas as a basis for determining the surface contributions to the tangential curvature.

From the results of the ray trace we calculate the tangential curvature of the image formed after refraction at successive surfaces of the system. Considering the rays *a* and *b* after refraction at surface *i*, a rotation of each ray so that their intersection point $\text{Ab}_i$ now lies in the plane of the *m* focus at a point distant $(\text{coma})'_{Ti}$ from the point $Prm_i$ would remove the curvature originally present in the image behind the surface *i*. The effect of such imagined rotations would be to introduce at the final image an amount of curvature opposite to that introduced by the first *i* surfaces of the system, that is, an amount $-\overset{i}{\Sigma}\, TXC'$. This leads, as before, to a measure of the surface contributions.

The angle through which the *q* ray must be rotated is

$$\begin{aligned} dU'_{ai} &= X'_{Ti}(\tan U'_{ai} - \tan U'_{pri}) \cos U'_{ai}/S'_{ai} \\ &= X'_{Ti}(\tan U'_{ai} - \tan U'_{pri}) \cos^2 U'_{ai}/(L_m' - X_a')_i, \end{aligned} \qquad \ldots\ldots(93)$$

with a corresponding expression for the rotation of the *b* ray given by

$$dU'_{bi} = X'_{Ti}(\tan U'_{pri} - \tan U'_{bi}) \cos^2 U'_{bi}/(L_m' - X_b')_i. \qquad \ldots\ldots(94)$$

The effect of such changes in the directions of the rays on the curvature of the final image is given by

$$dX_T' = dL'_{ab} = \frac{\partial L'_{ab}}{\partial U'_{ai}} dU'_{ai} + \frac{\partial L'_{ab}}{\partial U'_{bi}} dU'_{bi}, \qquad \ldots\ldots(95)$$

and then

$$\overset{i}{\Sigma}\, TXC' = -dL'_{ab} \qquad \ldots\ldots(96)$$

and

$$TXC_i' = \overset{i}{\Sigma}\, TXC' - \overset{i-1}{\Sigma}\, TXC'. \qquad \ldots\ldots(97)$$

The correction term to be applied to equation (95) is exactly similar to that for equation (85), already noted.

## REFERENCE

Conrady, A. E., 1929. *Applied Optics and Optical Design* (London: Oxford University Press).

# A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: IV. THE ESTIMATION OF THE TOLERANCES PERMISSIBLE IN THE PRODUCTION OF AN OPTICAL SYSTEM

By F. D. CRUICKSHANK,
University of Tasmania

*MS. received* 26 *February* 1945

*ABSTRACT.* A knowledge of the transfer coefficients for the aberrations permits the changes in the residual aberrations of the system to be calculated for any small departure from specified dimensions. Provided the changes in the residuals are small in comparison with the residuals themselves, the effect of the departure from the specified dimensions will be negligible. Hence the calculated values of the transfer coefficients provide a basis for the estimation of the limits within which each dimension of an optical system must be controlled in the workroom if systems having uniform characteristics are to be produced.

## § 1. INTRODUCTION

In the production of an optical system, serious imperfections may be introduced by failure to achieve the curvatures and other dimensions specified in the design. On the other hand, the cost of production may be increased considerably by striving to attain the specified dimensions within limits which are unnecessarily fine. It is thus an important part of the designer's work to specify the limits within which each quantity must be controlled during production. It is proposed now to examine this problem in the light of the analysis developed in the preceding papers of this series, and it will be shown that a method of estimating tolerances of this kind may be obtained which links on logically and conveniently with the general method of final design which has been developed. It is to be emphasized that the tolerances under consideration are not those which are generally termed "optical tolerances", which have relation to the amounts of residual aberrations which may be permitted in a system which aims at a certain standard of definition in the image. The present considerations relate to the degree of control to be exercised in the optical shop in the production of a given system if a uniform product is to be produced.

The final design of a system always represents more or less of a compromise. It is characterized by certain residual aberrations which have been calculated during the final stages of the design, and which experience, or the performance of a carefully built prototype, shows to be compatible with satisfactory image definition. The whole success of known methods of design depends on the fact that the mathematical analysis of the aberrations gives a fairly reliable guide to the actual physical aberrations of the system when constructed. In the course of production, any variation from the specified dimensions will result in a variation in the actual residual aberrations, but production will be satisfactory provided these variations are small compared with the residual aberrations themselves. Hence it is reasonable to base a system of estimating tolerances on the calculation of the effect on the residuals of a departure from the specified value of each quantity in the system. The reliability of these tolerances will be of the same order as the reliability of the design methods as a whole. The system of transfer coefficients developed in Part I provides the machinery for the method, little additional computation being required.

## § 2. THE TOLERANCES FOR THE CURVATURE OF EACH SURFACE OF THE SYSTEM

In Part I it is shown that it is possible to calculate a transfer coefficient which measures the rate of change of any aberration with the curvature of any surface of the system. These coefficients are calculated normally for the purpose of the final differential correction of the system. They are now available for the further purpose of estimating the tolerances permissible during production.

If we use the symbol $A'$ to denote any of the tangential aberrations of the system, then the calculated values of $\partial A'/\partial c$ reveal the effect of small changes of curvature on this aberration. After a close scrutiny of these quantities, tolerances may be set by assigning a permissible curvature variation at each surface such that the sum over all the surfaces of the effects on the residual aberrations due to the occurrance of such variations shall not exceeed some specified fractions of the values of the residual aberrations. If a curvature tolerance, $\pm dc_i$, is selected at each surface in this way, the corresponding tolerance in the radius of curvature is $\pm r_i^2\, dc_i$. It is often convenient to express the tolerance in terms of the number of fringes across a surface of a certain diameter. It is easily shown that the difference in curvature, $dc$, between two spherical surfaces of curvatures $c$ and $c+dc$, which show $x$ fringes across a surface of diameter $2a$ when placed together, is given very closely by

$$dc = \frac{x\lambda}{a^2}(1-0{\cdot}5\, a^2c^2). \qquad \ldots\ldots(98)$$

Writing $\partial A'/\partial c_f$ for the change in the aberration $A'$ per fringe change in curvature, we have for any surface

$$\frac{\partial A'}{\partial c_f} = \frac{\partial A'}{\partial c}\cdot\frac{\lambda}{a^2}(1-0{\cdot}5\, a^2c^2). \qquad \ldots\ldots(99)$$

Equation (99) provides the necessary basis for the estimation of the curvature tolerances.

## § 3. THE TOLERANCE IN THE AXIAL THICKNESS OF A COMPONENT OR AIR SPACE

The thickness of a lens is a quantity which is much more difficult to control in production than the curvatures of its surfaces, so that the question of tolerances becomes very important. For any component the transfer coefficient, $\partial A'/\partial d$, for its rear surface specifies the change of the aberration per unit change of axial thickness. In accordance with the signs we have used, a negative $d$-change increases the thickness of the component. Formally, then, we introduce an axial thickness coefficient, $\partial A'/\partial t$, defined by

$$\frac{\partial A'}{\partial t_h} = -\left(\frac{\partial A'}{\partial d_2}\right)_h, \qquad \ldots\ldots(100)$$

the subscript $h$ referring to the general component and the subscript 2 denoting the second, or rear, surface of that component. Thus, without further computation, a set of coefficients is available for the estimation of the tolerances for the thicknesses of the components and air spaces. We make use of them by assigning to each component and air space a permissible thickness variation such that the sum over all the components of the effects on the residual aberrations due to such variations shall not exceed some small fraction of the measures of the residual aberrations.

## §4. THE TOLERANCES FOR THE REFRACTIVE INDEX AND DISPERSION OF THE GLASSES

It was shown in Part I that for any monochromatic aberration, $A'$, there may be calculated a differential coefficient, $\partial A'/\partial N_h$, which measures the change of this aberration in the image formed by the system per unit change of refractive index of the component, $h$, of the system. These coefficients provide full information as to the limits within which the refractive index of each component must be controlled for satisfactory production of the system, and permit a system of tolerances to be established. As before, we assign a permissible variation of the refractive index to each component such that the sum over all the components of the effects on the residual aberrations due to the permitted variations is less than a prescribed small fraction of these residuals. As regards the variation of the dispersions of the glasses, the coefficients, $\partial Lch'/\partial P_h$ and $\partial Tch'/\partial P_h$, developed in Part I furnish the basis for tolerances, but generally speaking, the variation of the dispersion of a glass between successive melts is negligible, so that these tolerances are seldom required. Frequently a more serious effect of the variation of refractive index is the resulting change in focal length of the system. This is dealt with in the next section.

## §5. THE TOLERANCES FOR THE CONTROL OF THE FOCAL LENGTH OF THE SYSTEM

It is often important to ensure that the focal length of a particular system is controlled within fine limits during production. For this we must know the effect of the variation of curvature, thickness, and refractive index on the focal length. From the results of the paraxial ray trace

$$f' = y/u_k'$$

from a ray incident parallel to the axis at a height $y$ on the first surface. Hence

$$\begin{aligned}\frac{\partial f'}{dc_i} &= -(y/u_k'^2)\frac{\partial u_k'}{\partial c_i}\\ &= -(f'/u_k')\frac{\partial u_k'}{\partial c_i}. \qquad \ldots\ldots(101)\end{aligned}$$

Corresponding expressions hold for $\partial f'/\partial n_i$ and $\partial f'/\partial d_i$. For the effect of the variation in the thickness of a component we have

$$\frac{\partial f'}{\partial t_h} = (f'/u_k')\left(\frac{\partial u_k'}{\partial d_2}\right)_h, \qquad \ldots\ldots(102)$$

and the effect of refractive-index variation is given by

$$\frac{\partial f'}{\partial N_h} = -\frac{\partial f'}{\partial n_1}\frac{N_{h-1}}{N_h^2} + \frac{\partial f'}{\partial n_2}\frac{1}{N_{h+1}}. \qquad \ldots\ldots(103)$$

The transfer coefficients in equations (101) to (103) are quickly computed as the values of $\partial u_k'/\partial c_i$, etc., have been calculated in the general computation for the paraxial ray. Thus a series of tolerances may be set quite simply for the control of the focal length of the system.

## §6. THE USE OF COMPONENTS OUTSIDE THE SPECIFIED TOLERANCES

When a large number of optical systems of a certain type are being produced, a certain percentage of rejected components with dimensions outside the specified tolerances is inevitable. The intelligent use of the information provided by the transfer coefficients calculated during the design enables some at least of these rejects to be sorted into sets of optics which will give a satisfactory performance. Components rejected on account of non-spherical figure are excluded from our considerations.

The rejected components are classified according to the type and amount of their departure from specified dimensions. For each aberration a limit is set for the additional amount of this aberration which can be admitted on account of departure from exact specifications. This will be some small fraction of the calculated residual aberration. A table is prepared which sets out the aberration changes for departure from the specified dimensions in terms of the units used in the workroom. The trained worker then derives from this table the amount of each aberration introduced into the system if a certain reject component, A, is employed. A study of the transfer coefficients in the table will now reveal whether some departure from dimensions of another component, B, will introduce amounts of the various aberrations which will compensate those introduced by A, and thus bring their totals within the limits prescribed. If this is possible, a reject of type A is paired with B, etc., and in this way use may be made of the rejected components. This kind of attempt to salvage components is probably made in every optical shop by trial-and-error methods or by bench tests. The use of a table of coefficients as suggested organizes such attempts intelligently, giving a reliable guide to what is possible in this respect.

# A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: V. TRANSFER COEFFICIENTS FOR THE ASTIGMATISM AT SMALL APERTURE AND FINITE OBLIQUITY

By F. D. CRUICKSHANK,
University of Tasmania

*MS. received* 26 *February* 1945

*ABSTRACT.* Transfer coefficients are developed which specify the changes of the positions of the tangential and sagittal foci in narrow pencils of any obliquity due to small changes in curvature, refractive index, and axial separation which may be made within the system.

---

## §1. INTRODUCTION

In the preceding papers of this series, transfer coefficients have been developed for the tangential aberrations only of a lens system. To complete the development it is necessary to establish a means of handling the astigmatism in a similar manner. The full solution of this problem, however, requires the extension of the foregoing methods to the general skew trace. Not having had opportunity to give much attention to this problem as yet, the writer has used a simple analysis of the astigmatism at small aperture in pencils of any obliquity, checking the astigmatism at full aperture from time to time by the usual skew trace. This constitutes quite a useful designing tool, as the changes in the astigmatism at small aperture due to changes within the system can be estimated for the differential correction process. The transfer coefficients for the shifts of the foci of narrow fans of tangential and sagittal rays are deduced, and this provides a basis for estimating the astigmatism changes produced by any alteration made within the system.

## §2. THE TRANSFER COEFFICIENTS FOR THE SHIFT OF THE TANGENTIAL FOCUS

In addition to the principal ray of the oblique pencil of given obliquity, we trace another ray of the same obliquity close to the principal ray and calculate for it the single surface coefficients and the transfer coefficients in the normal way. We denote this ray by the subscript *cp*, the letters standing for *close principal* ray. From the trace of the principal ray we have, as in Part I, equation (45 *a*),

$$t_k' = \frac{\partial p_k'}{\partial p_1} \bigg/ \frac{\partial U_k'}{\partial p_1}.$$

The two rays *pr* and *cp* will emerge from the last surface of the system at different points, and it is obvious that we may specify the separation of the emergence points by

$$\delta p'_{prk} = t_k'(U'_{cp} - U'_{pr})_k = t_k'\,\delta U_k'. \qquad \ldots\ldots(104)$$

If a change is now made at some surface within the system, the two traced rays will emerge from points with a different separation and will intersect in the final image-space under a new angle. Thus if we can calculate the changes in $\delta p'_{prk}$ and $\delta U_k'$ resulting from the alteration made within the system, we can calculate the new value of $t_k'$ specifying the position of the tangential focus.

Let us suppose that a small curvature change is to be made at surface *i* within the system. Differentiating equation (104),

$$\frac{\partial}{\partial c_i}(\delta p'_{prk}) = t_k'\frac{\partial}{\partial c_i}(U'_{cp} - U'_{pr})_k + \frac{\partial t_k'}{\partial c_i}(U'_{cp} - U'_{pr})_k.$$

For the term on the left-hand side of this equation we have

$$\frac{\partial}{\partial c_i}(\delta p'_{prk}) = \frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i},$$

and combining these two equations we obtain

$$\frac{\partial t_k'}{\partial c_i} = \frac{1}{\delta U_k'}\left[\left(\frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i}\right) - t_k'\left(\frac{\partial U'_{cpk}}{\partial c_i} - \frac{\partial U'_{prk}}{\partial c_i}\right)\right]. \qquad \ldots\ldots(105)$$

The corresponding expressions for thickness changes and refractive-index changes may be written down, giving

$$\frac{\partial t_k'}{\partial d_i} = \frac{1}{\delta U_k'}\left[\left(\frac{\partial p'_{cpk}}{\partial d_i} - \frac{\partial p'_{prk}}{\partial d_i}\right) - t_k'\left(\frac{\partial U'_{cpk}}{\partial d_i} - \frac{\partial U'_{prk}}{\partial d_i}\right)\right], \qquad \ldots\ldots(106)$$

$$\frac{\partial t_k'}{\partial n_i} = \frac{1}{\delta U_k'}\left[\left(\frac{\partial p'_{cpk}}{\partial n_i} - \frac{\partial p'_{prk}}{\partial n_i}\right) - t_k'\left(\frac{\partial U'_{cpk}}{\partial n_i} - \frac{\partial U'_{prk}}{\partial n_i}\right)\right]. \qquad \ldots\ldots(107)$$

It will thus be seen that the transfer coefficients for the *t*-focus require little calculation beyond the work involved in the trace and coefficients of the ray *cp*. If changes are made during the process of differential correction which are too large to be treated as differentials, an improvement in accuracy may be obtained by calculating the new values of $\delta p_k'$ and $\delta U_k'$ resulting from the change and deducing the new value of $t_k'$ directly from them. Thus, for a curvature change $\delta c_i$, we have

$$\text{new}\,(\delta p_k') = \delta p_k' + \left(\frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i}\right)\delta c_i, \qquad \ldots\ldots(108)$$

$$\text{new}\,(\delta U_k') = \delta U_k' + \left(\frac{\partial U'_{cpk}}{\partial c_i} - \frac{\partial U'_{prk}}{\partial c_i}\right)\delta c_i, \qquad \ldots\ldots(109)$$

and the change in the *t*-focus is given by

$$\delta t_k' = \frac{(\delta p_k')_{\text{new}}}{(\delta U_k')_{\text{new}}} - t_k'. \qquad \ldots\ldots(110)$$

## § 3. THE EFFECT OF CHANGES AT A SINGLE SPHERICAL SURFACE ON THE FOCUS OF A NARROW FAN OF SAGITTAL RAYS

The position of the sagittal focus behind the successive surfaces of a system is easily calculated by the well-known *s*-trace which is associated with the ordinary trace of the principal ray of the oblique pencil under consideration. We propose to use the normal relations of the *s*-trace to develop the transfer coefficients for the sagittal focus. The position of the sagittal focus behind any surface of the system is given by

$$\frac{N'}{s'}=\frac{N'\cos I'-N\cos I}{r}+\frac{N}{s}. \qquad \ldots\ldots(111)$$

If we now write $\sigma=1/s,\quad \sigma'=1/s,$ ......(112)

and $G=(N'\cos I'-N\cos I)/N',$ ......(113)

equation (111) becomes

$$\sigma'=Gc+n\sigma. \qquad \ldots\ldots(114)$$

This procedure clears the basic equation of reciprocals and makes it easier to handle in the subsequent differentiations. Let us suppose now that a small curvature change, $dc_i$, is to be made at the surface *i* of the system. The effect on the *s*-focus behind the surface will be determined by the derivative

$$\frac{\partial\sigma_i'}{\partial c_i}=G_i+c_i\frac{\partial G_i}{\partial c_i}+n_i\frac{\partial\sigma_i}{\partial c_i}.$$

The change $\sigma_i$ due to a small change of curvature is of second order only, so that we may neglect the last term on the right-hand side of the above equation, and hence

$$\frac{\partial\sigma_i'}{\partial c_i}=G_i+c_i\frac{\partial G_i}{\partial c_i}. \qquad \ldots\ldots(115)$$

To be able to compute the derivative $\partial\sigma_i'/\partial c_i$ we require an expression for $\partial G_i/\partial c$. This may be obtained by differentiation of equation (113) and the refraction equation $N\sin I=N'\sin I'$. Thus,

$$\begin{aligned}\frac{\partial G}{\partial I}&=-\sin I'\frac{\partial I'}{\partial I}+n\sin I\\&=-\sin I'\,n\frac{\cos I}{\cos I'}+n\sin I\\&=\sin I'\left(1-n\frac{\cos I}{\cos I'}\right)\\&=\sin I'\left(1-\frac{\partial U'}{\partial U}\right).\end{aligned} \qquad \ldots\ldots(116)$$

Differentiation of the computing equation

$$(L-r)\sin U=r\sin I$$

gives $\dfrac{\partial I}{\partial c}=L\sin U/\cos I,$

and hence we have

$$\begin{aligned}\frac{\partial G}{\partial c}&=\frac{\partial G}{\partial I}\frac{\partial I}{\partial c}\\&=\sin I'\left(-1\frac{\partial U'}{\partial U}\right)\frac{L\sin U}{\cos I}\\&=\sin I'\frac{\partial U'}{\partial c}.\end{aligned} \qquad \ldots\ldots(117)$$

The computation of $\partial G/\partial c$ at any surface thus involves only one operation beyond the calculation of the ordinary single surface coefficients, which will have been computed already for the principal ray. Two further operations complete the calculation of $\partial \sigma'/\partial c$.

To determine the effect of a small change in refractive index at any surface $i$ of the system, differentiation of equation (114) gives

$$\frac{\partial \sigma_i'}{\partial n_i} = \sigma_i + c_i \frac{\partial G}{\partial n_i}, \qquad \ldots\ldots(118)$$

in which, by differentiation of (113), we have

$$\begin{aligned} \frac{\partial G}{\partial n} &= -\sin I' \frac{\partial I'}{\partial n} - \cos I \\ &= -\sin I' \frac{\sin I'}{\cos I'} - \cos I \\ &= \sin I' \frac{\partial U'}{\partial n} - \cos I. \end{aligned} \qquad \ldots\ldots(119)$$

In this case only four operations are required to calculate $\partial \sigma'/\partial n$ for any surface after the single surface coefficients.

Finally, if a small change is made in the axial separation of two consecutive surfaces of the system, differentiation of equation (114) gives

$$\frac{\partial \sigma_i'}{\partial d_i} = c_i \frac{\partial G_i}{\partial d} + n_i \frac{\partial \sigma_i}{\partial d_i}. \qquad \ldots\ldots(120)$$

Expressions for the two derivatives on the right-hand side of this equation may be deduced quite easily. Thus, differentiating equation (113),

$$\begin{aligned} \frac{\partial G}{\partial d} &= -\sin I' \frac{\partial I'}{\partial d} + n \sin I \frac{\partial I}{\partial d} \\ &= \sin I' \left( \frac{\partial I}{\partial d} - \frac{\partial I'}{\partial d} \right) \\ &= \sin I' \left( \frac{\partial I}{\partial d} - n \frac{\cos I}{\cos I'} \frac{\partial I}{\partial d} \right) \\ &= \sin I' \left( 1 - \frac{\partial U'}{\partial U} \right) \frac{\partial I}{\partial p} \frac{\partial p}{\partial d} \\ &= \sin I' \sin U \frac{\partial U'}{\partial p}. \end{aligned} \qquad \ldots\ldots(121)$$

To obtain an expression for the second derivative we differentiate the transfer formula of the $s$-trace, which is

$$s_i = s_{i-1}' - D_{i-1}', \qquad \ldots\ldots(122)$$

and thus

$$\begin{aligned} \frac{\partial \sigma_i}{\partial d_i} &= -\frac{1}{s_i^2} \frac{\partial s_i}{\partial d_i} \\ &= \sigma_i^2 \frac{\partial D_{i-1}'}{\partial d_i} \\ &= -\sigma_i^2 \sec U_i \left[ 1 - \frac{\sin U_i \sin (U+I)_i}{\cos I_i} \right]. \end{aligned} \qquad \ldots\ldots(123)$$

The last step follows simply from the well known relation between $D_{i-1}'$ and $d_i$.

Often the bracketed expression may be replaced by unity with a good degree of approximation, and is accurately unity for a plane surface. Equations (120), (121) and (123) permit the calculation of $\partial\sigma'/\partial d$ at any surface of the system.

## §4. THE TRANSFER COEFFICIENTS FOR THE SAGITTAL FOCUS OF A NARROW PENCIL

It now remains to develop a method of computing the transfer coefficients for $\sigma'$-changes so that the effect of a change within the system on the final value of $\sigma_k'$ may be determined. Since $G$ is a function of the angles of incidence and refraction, its value at any surface will change with change of the point of incidence. It will be found, however, that in general the change of $G$ along the path of the pencil due to small alterations within the system has a very small effect on the transfer coefficients, and, on account of the simplification obtained by neglecting this factor, it is proposed to develop a set of transfer coefficients which leave it out of consideration altogether.

Suppose that by some means a change $d\sigma_i'$ is produced in the narrow pencil refracted at surface $i$. To determine the effect of this change at the following surface of the system we require an expression for the derivative $\partial\sigma_{i+1}/\partial\sigma_i'$. We may write the transfer formula in (122) in the form

$$\sigma_{i+1} = \sigma_i'(1 - D_i'\sigma_i')^{-1}, \qquad \ldots\ldots(124)$$

and hence

$$\frac{\partial\sigma_{i+1}}{\partial\sigma_i'} = (1 - D_i'\sigma_i')^{-2}. \qquad \ldots\ldots(125)$$

The change which will result behind surface $(i+1)$ must now be determined. From equation (114) we have

$$\frac{\partial\sigma'_{i+1}}{\partial\sigma_{i+1}} = n_{i+1}, \qquad \ldots\ldots(126)$$

whence

$$\frac{\partial\sigma'_{i+1}}{\partial\sigma_i'} = n_{i+1}\frac{\partial\sigma_{i+1}}{\partial\sigma_i'},$$

and extending this expression surface by surface we obtain

$$\frac{\partial\sigma_k'}{\partial\sigma_i'} = \frac{\partial\sigma_{i+1}}{\partial\sigma_i'}n_{i+1}\frac{\partial\sigma_{i+2}}{\partial\sigma'_{i+1}}n_{i+2}\cdots\cdots\frac{\partial\sigma_k}{\partial\sigma'_{k-1}}n_k, \qquad \ldots\ldots(127)$$

and the relation between these transfer coefficients at successive surfaces is, therefore,

$$\frac{\partial\sigma_k'}{\partial\sigma_i'} = \frac{\partial\sigma_{i+1}}{\partial\sigma_i'}n_{i+1}\cdot\frac{\partial\sigma_k'}{\partial\sigma'_{i+1}}. \qquad \ldots\ldots(128)$$

We now have the complete framework for the computation of the transfer coefficients. Equation (125) is used first to calculate $\partial\sigma_{i+1}/\partial\sigma_i'$ for each surface except the last. Then at the last surface $\partial\sigma_k'/\partial\sigma_k' = 1$, and equation (128) is used to calculate $\partial\sigma_k'/\partial\sigma_{k-1}'$, and the computation may be continued surface by surface through the system. Finally these basic transfer coefficients are used in conjunction with the single surface coefficients, giving expressions of the form

$$\frac{\partial\sigma_k'}{\partial c_i} = \frac{\partial\sigma_k'}{\partial\sigma_i'}\,\frac{\partial\sigma_i'}{\partial c_i},$$

and returning to the quantity $s'$ instead of its reciprocal, we have

$$\frac{\partial s_k'}{\partial c_i} = -\,s_k'^2\frac{\partial\sigma_k'}{\partial\sigma_i'}\,\frac{\partial\sigma_i'}{\partial c_i}, \qquad \ldots\ldots(129)$$

with corresponding expressions for $\partial s_k'/\partial d_i$ and $\partial s_k'/\partial n_i$.

## §5. A NUMERICAL EXAMPLE

In the course of correcting a certain photographic objective having ten surfaces, a change of curvature of 0·001 mm.$^{-1}$ was made at the seventh surface, the radius of curvature of that surface changing from 1091 mm. to 521·78 mm. In a pencil of approximately 17° obliquity the following results were obtained:

$$\frac{\partial s_k'}{\partial c_7} = -2112{\cdot}5, \qquad \frac{\partial t_k'}{\partial c_7} = -2416{\cdot}0,$$

and hence

$$ds_k' = -2{\cdot}113, \qquad dt_k' = -2{\cdot}416,$$

giving

$$\text{new } s_k' = 47{\cdot}273, \qquad \text{new } t_k' = 47{\cdot}661.$$

Using the alternative relations (108) to (110) there was obtained

$$dt_k' = -2{\cdot}335,$$
$$\text{new } t_k' = 47{\cdot}742.$$

A full trace of the altered system gave the final positions of the two foci as

$$s_k' = 47{\cdot}338, \qquad t_k' = 47{\cdot}801,$$

which represents an accuracy of about 3%.

---

# A TRANSFER METHOD FOR DERIVING THE EFFECT ON THE IMAGE FORMED BY AN OPTICAL SYSTEM FROM RAY CHANGES PRODUCED AT A GIVEN SURFACE

By A. L. M'AULAY,
University of Tasmania

*MS. received* 25 *January* 1944

*ABSTRACT.* A transfer method is described depending on the invariance of a function of the position change of successive foci of a narrow pencil as it passes through an optical system.

The method described is approximate, as the invariant is applied to wide pencils, but by reason of its speed it is a useful tool in design.

## §1. INTRODUCTION

In a paper by M'Aulay and Cruickshank (1945) a method is described by which the effect on the final image formed by an optical system may be computed when alterations are made at any surface of the system. The method naturally breaks into two parts. In the first part the changes which rays suffer as they pass through the surface under consideration are evaluated, and in the second part these changes are transferred to the final image formed by the system.

The present paper describes another method of transferring the changes produced at any surface to the final image. It is based on a theorem of a form similar to Lagrange's, but is applicable to pencils of any obliquity.* If the theorem is applied to narrow pencils it is exact for differential changes, but in the use here described it is applied to wide pencils and is approximate. It has been used principally as a primary computing method where speed is more important than exactness, and it has been found of considerable practical value. It is believed that its potentialities are greater than the particular development discussed.

## § 2. AN INVARIANT OF THE LAGRANGE FORM FOR OBLIQUE PENCILS

The method deals essentially with pencils, not with rays. It follows the change in position of the successive foci of a pencil which suffers a small change in position due to an alteration made in the constants of one surface of the system through which it is passing. Figure 1 shows a narrow pencil ABC focusing at B after passing through the *i*th surface of the system. An alteration made to the *i*th surface causes the pencil to focus instead at B.* The co-ordinates of the change are $w_j$ and $z_j$ measured along and at right angles to the original axis of the pencil. They are given the positive sign when they have the direction shown in the figure, as this relates them conveniently to the usual sign of $U$, the angle under an incident ray. If $\gamma_j$ is the angle between the extreme rays of the pencil and $N_j$ is the refractive index of the medium between the *i*th and the *j*th surfaces, it may be shown that

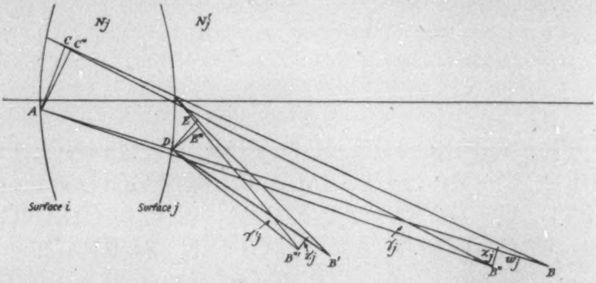


Figure 1.

(i) $N\gamma z$ remains unchanged as the pencil is followed through successive surfaces in its original and changed form, and

(ii) $N\gamma^2 w$ also remains unchanged.

The second relation holds, however, only if the change is such that a spherical wave remains spherical after the change. It is a useful relation for some purposes, but will not be further considered here.

---

* This theorem, given in § 2, was discovered by the author. The referee has pointed out that it was known independently (see, e.g., v. Rohr, *Formation of Images in Optical Instruments*, 1920, p. 174), and no proof is therefore given here.

It is important to note that the change of a ray is not being followed. The $z$s are co-ordinates of the changes in the successive focal points of a pencil, but the actual rays which have been followed in the above argument change from surface to surface. Unless this point is kept clearly in mind it is easy to make serious mistakes when the principle is used as a computing tool.

*Alternative statement of the principle.* Figure 2 shows a pencil with its axis displaced a small distance $z'$. It is evident that if $\alpha'$ is the angle through which the axis is displaced,

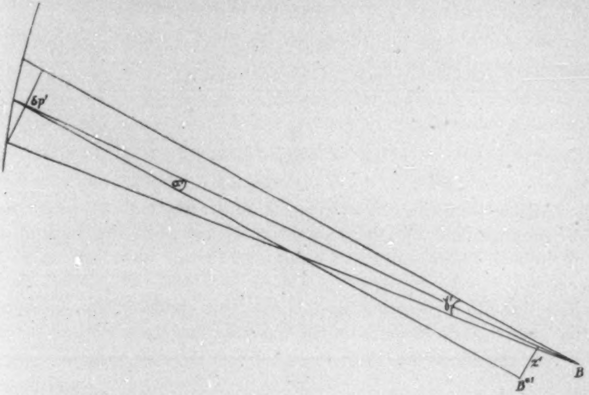


Figure 2.

$$N'\gamma' z' = N'\frac{\delta p'}{S'} z' = N'\alpha'\delta p',$$

where $S'$ is the length of the axis of the pencil from surface to focus, $\delta p'$ is the width of the pencil at the surface and $\alpha'$ the angle turned through by the axis of the pencil due to the alteration. This form is often best suited for obtaining the value of the invariant at an altered surface.

## §3. APPLICATION OF THE PRINCIPLE TO THE TRANSFER OF DIFFERENTIAL CHANGES PRODUCED AT THE $i$TH SURFACE OF A SYSTEM

In the paper by M'Aulay and Cruickshank quoted above, a change in the aberrations of an optical system due to an alteration of the $i$th surface was discussed. The system is primarily analysed by considering six rays, three axial (marginal, zonal and paraxial) and three oblique ($a$, $pr$ and $b$). The same rays are considered in the method to be described, but they are treated as the axes of six pencils. It is shown in the paper quoted that the change in angle under a ray refracted at a spherical surface can be given as a function of the curvature or refractive-index change. In what follows, it will be assumed that such a surface alteration has produced changes $dU'$ in the angles under the refracted rays, and that these have been computed.

The transfer theorem is used to convert these $dU$ to displacements $z'$ perpendicular to the directions of the changed rays in the neighbourhood of the final image plane. It is evident that $z'$ is very nearly the co-ordinate of the displacement of the focus of the narrow pencil which in the notation of the last section would have been written $z_k'$, where $k$ is the final surface.

*Approximate evaluation of $z'$.* The theorem is used in the form

$N'\alpha'\delta p'$ is invariant through the system.

This invariant is found for the $i$th surface and used to obtain $z'$ after refraction at the last surface by the relation

$$N_i'\alpha_i'\delta p_i' = N_k'\alpha_k'\delta p_k' = N_k'\frac{z'}{S'}\delta p_k',$$

or

$$z' = N_i'\alpha_i'\frac{\delta p_i'}{\delta p_k'}S_k',$$

where the refractive index after the last surface considered is unity.

$N_i'$ is available from the original data, $\alpha'$ is the change $dU_i'$ which has been supposed computed, and $S_k'$, the distance measured along the ray from the last surface to the final image, can easily be obtained from the trace. It remains to evaluate $\delta p_i'/\delta p_k'$.

It is here that the most serious approximation is made. The object is to find sufficiently good values of $\delta p'$ from the basic six-ray trace without much extra computation. In the tracing method used, the value of $L'\sin U' = p'$ is, incidentally, found for each ray. This is the perpendicular distance from the pole of a surface to the ray under consideration.

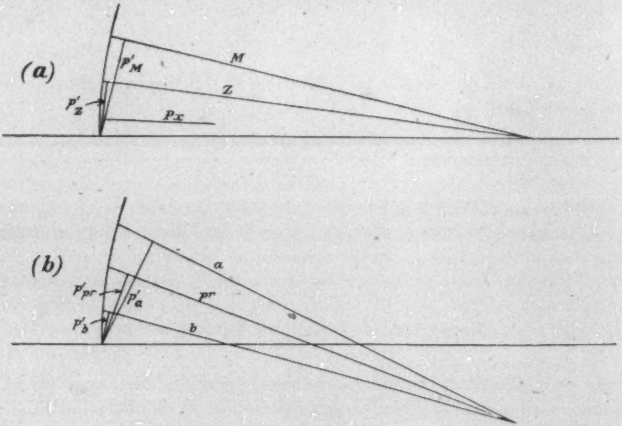


Figure 3.

Figure 3 (a) shows marginal, zonal and paraxial rays; figure 3 (b), *a*, *pr* and *b* rays and the perpendiculars dropped from the pole on them. The following values are taken for $(\delta p'/\delta p_k')$:

$$\begin{aligned}
\delta p_x'/\delta p'_{xk} &= p_x'/p'_{xk},\\
\delta p_z'/\delta p'_{zk} &= p_M'/p'_{Mk},\\
\delta p_M'\delta p'_{Mk} &= 2(p_M'-p_Z')/(p_M'-p_Z')_k - p_Z'/p'_{Zk},\\
\delta p'_{pr}/\delta p'_{prk} &= (p_a'-p_b')/(p_a'-p_b')_k,\\
\delta p_a'/\delta p'_{ak} &= (p_a'-p'_{pr})/(p_a'-p'_{pr})_k\\
&\quad + \tfrac{1}{2}[(p_a'-p'_{pr})/(p_a'-p'_{pr})_k - (p'_{pr}-p_b')/(p'_{pr}-p_b')_k],\\
\delta p_b'/\delta p'_{bk} &= (p'_{pr}-p_b')/(p'_{pr}-p_b')_k\\
&\quad - \tfrac{1}{2}[(p_a'-p'_{pr})/(p_a'-p'_{pr})_k - (p'_{pr}-p_b')/(p'_{pr}-p_b')_k].
\end{aligned}$$

It is evident that other extrapolations for the marginal, *a* and *b* rays might be used. The rather crude forms given are easy to compute and reasonably good.

## §4. THE ABERRATION CHANGES FROM THE $z$

*Spherical aberration and curvature.* The aberration changes can be stated simply in terms of $z'$. The changes in spherical aberration and tangential curvature of field will be written down without further discussion as their derivation follows the same lines as that given by Cruickshank, Part I (1945).

For spherical aberration

$$dL_M' = dl' - dL_M' = z_x'/u' - z_M' \operatorname{cosec} U_M',$$

with an exactly similar expression for zonal spherical aberration.

For curvature

$$dX_T' = (z_a' \cos U_b' - z_b' \cos U_a') \operatorname{cosec}(U_a' - U_b') - z_s' \operatorname{cosec} U_s'.$$

The notation is that of Conrady's *Applied Optics and Optical Design* or the papers quoted above. A change of sign, however, has been introduced for $X_T'$, and the signs for coma and distortion will also be changed. It has been found convenient in this work to adopt a positive sign for an aberration that arises when parallel pencils of small inclination to the principal axis pass from air through a convex glass surface.

*Coma and distortion.* The co-ordinate changes perpendicular to the principal rays can be used direct to estimate these aberrations. The change in coma is sufficiently well given by the change in distance between the principal ray and the median of the *a* and *b* rays. It is given by

$$d(\text{coma}') = z'_{pr} - \tfrac{1}{2}(z_a - z_b')\sec\tfrac{1}{2}(U_a' - U_b').$$

The treatment of change of distortion, involving as it does an imaginary ray not included in the trace, requires a little more discussion. An ideal ray is considered which is identical with the principal ray till it strikes the surface at which an alteration takes place. Subsequently this ray is treated as though it were refracted according to the paraxial form of the computing equations. The change in angle under the ideal ray after refraction at the altered surface therefore differs for that obtained for the principal ray. This change of angle is computed and is transferred to a $z'$ change in the neighbourhood of the final image, which is written $z'_{prx}$ and is given by

$$z'_{prx} = N'\alpha'_{prx}(p_x'/p'_{xk})l_k'.$$

The change in distortion is now given in the same manner as the change in coma by

$$d(\mathrm{dist}') = z'_{prx} - z'_{pr}.$$

REFERENCES

M'Aulay, A. L. and Cruickshank, F. D., 1945. *Proc. Phys. Soc.* **57**, 302.
Cruickshank, F. D., 1945. *Proc. Phys. Soc.* **57**, 350.

---

# THE RELATION BETWEEN THE BRIGHTNESS AND TEMPERATURE OF A TOTAL RADIATOR

By E. F. CALDIN,
Queen's College, Oxford

*MS. received* 31 *January* 1945 ; *in revised form* 14 *June* 1945

*ABSTRACT.* Values are calculated for the brightness of a total radiator at temperatures between 2000° K. and 3120° K., using the standard C.I.E. relative luminosity curve and the Planck radiation law. A simple logarithmic equation is found to express the relation between brightness and temperature to within ±1% over this temperature range. The ratio of the new international candle to the lightwatt is calculated from the results.

## §1. INTRODUCTION

It is sometimes desirable to use an expression for the brightness of a total radiator as a function of temperature. From the literature it appears that nothing has been published on the subject since the adoption of the C.I.E. relative luminosity curve in 1924. Hyde, Forsythe and Cady (1919) published some values of the brightness for the range 1700° K. to 2650° K.; they used their own values for the relative luminosity function, however, and these differ from the standard values later adopted. They expressed the results in terms of a rather complex four-term equation. Nernst (1906) expressed the relation experimentally found between the brightness ($B$) and temperature ($T$) of a total radiator for the range 1460° K. to 2280° K. by the following equation:—

$$\log B = -A/T + C,$$

but the relative luminosity curve for the observer concerned is not recorded.

Values for the brightness of a total radiator have been calculated from the C.I.E. relative luminosity data and the Planck equation for fifteen temperatures ranging from 2000° K. to 3120° K. The following approximate formula, which has the same form as Nernst's relation, has been obtained for the brightness of a total radiator over this temperature-range :—

$$\log_{10} B = 4{\cdot}275 - 10860/T, \qquad \ldots\ldots(1)$$

where $B$ is the brightness of the total radiator (lightwatts cm.$^{-2}$) and $T$ is its temperature (°K.).

## §2. CALCULATION OF BRIGHTNESS

The brightness of a total radiator at a given temperature is calculated from the equation

$$B = \frac{1}{\pi}\int_0^\infty J_\lambda V_\lambda d\lambda, \qquad \ldots\ldots (2)$$

where $B$ is the brightness (lightwatts cm.$^{-2}$), $J_\lambda d\lambda$ is the total energy flux per unit area from a total radiator (watts cm.$^{-2}$) between the wave-lengths $\lambda \pm \frac{d\lambda}{2}$ (microns), and $V_\lambda$ is the value of the relative luminosity factor at the wave-length $\lambda$, as given by the 1924 C.I.E. standard luminosity curve (Judd, 1931). The factor $\frac{1}{\pi}$ appears because the brightness of an element of surface, i.e. the ratio of intensity to projected area in a given direction, bears the ratio $\frac{1}{\pi}$ to the total luminous flux per unit area from the same element of surface (assuming that the surface has the properties of a perfect diffuser). In the present case the expression $\int_0^\infty J_\lambda V_\lambda d\lambda$ represents the total luminous flux per unit area of surface, since it is derived from $J_\lambda d\lambda$, which represents the total energy flux per unit area for a small wave-length range. Hence the brightness is given by $\frac{1}{\pi}\int_0^\infty J_\lambda V_\lambda d\lambda$.

The values of $J_\lambda d\lambda$ are derived from Planck's formula,

$$J_\lambda d\lambda = \frac{C_1 \lambda^{-5} d\lambda}{e^{C_2/\lambda T} - 1}, \qquad \ldots\ldots (3)$$

where $C_1$ and $C_2$ are constants and $J_\lambda d\lambda$, $\lambda$, and $T$ have been previously defined.

Convenient tables for deducing values of $B$ from equation (2) are those of Skogland (1929). These tables employ the values for $C_1$ and $C_2$ given in the International Critical Tables, namely $C_1 = 3{\cdot}703 \times 10^{-5}$ erg sec.$^{-1}$ cm.$^{+2}$, $C_2 = 14330$ micron degrees. The range of temperature considered is 2000° K. to 3120° K.; this is covered in steps of 20° in table 1 and 80° in table 2. Table 1 gives at each temperature, besides values of $J_\lambda/J_{0{\cdot}59}$ for a series of wave-lengths, values of $J_{0{\cdot}59}/C_1$ to 5 significant figures (units: microns$^{-5}$). From these latter values, taking $C_1 = 3{\cdot}703 \times 10^{-5}$ erg sec.$^{-1}$ cm.$^{+2}$ and $V_{0{\cdot}59} = 0{\cdot}7570$ (Judd, 1931), we calculate the value of $J_{0{\cdot}59}V_{0{\cdot}59}$ at a series of temperatures, at intervals of 80°, from 2000° K. to 3120° K. Skogland's table 2 gives at each of these temperatures, for a series of wave-lengths from $0{\cdot}40\,\mu$ to $0{\cdot}76\,\mu$ in steps of $0{\cdot}01\,\mu$, values for $J_\lambda V_\lambda/(JV)_{\max}$, where $(JV)_{\max}$ is the maximum value of $J_\lambda V_\lambda$ at the given temperature. Reading off the values for $J_{0{\cdot}59}V_{0{\cdot}59}/(JV)_{\max}$, and substituting the value already obtained for $J_{0{\cdot}59}V_{0{\cdot}59}$, we find $(JV)_{\max}$ for the given temperature. Table 2 gives also the sum of the values of $J_\lambda V_\lambda/(JV)_{\max}$, at each temperature, i.e.,

$$\sum_{\lambda=0{\cdot}40\,\mu}^{\lambda=0{\cdot}76\,\mu} J_\lambda V_\lambda/(JV)_{\max},$$

the summation referring to steps of $0{\cdot}01\,\mu$. The values at the wave-length limits are not more than about 0·03% of the maximum values, and the sum does not differ by more than 1 in $10^4$ from

$$\sum_{\lambda=0}^{\lambda=\infty} J_\lambda V_\lambda/(JV)_{\max}.$$

Multiplying the sum by 0·01, we obtain, therefore, a value (in microns) of

$$\int_0^\infty [J_\lambda V_\lambda/(JV)_{\max}]\,d\lambda, \quad \text{i.e.}\ \frac{1}{(JV)_{\max}}\int_0^\infty J_\lambda V_\lambda\,d\lambda,$$

which is accurate at least within 1 in $10^4$. Multiplying this by the value of $(JV)_{\max}$ already found, we obtain the value of $\int_0^\infty J_\lambda V_\lambda\,d\lambda$; and hence, dividing by $\pi$, the brightness of a total radiator, at the given temperature.

## § 3. RESULTS AND DISCUSSION

In the table below are given values for the brightness $B$, in lightwatts per sq. cm., and for $\log_{10} B$, at a series of temperatures from 2000° K. to 3120° K.

The value of $\log_{10} B$ is found to vary nearly linearly with $1/T$, and the following formula has been tested :—

$$\log_{10} B = 4{\cdot}2750 - 10860{\cdot}0/T. \qquad \ldots\ldots(1)$$

Values derived from this formula are given in the table and compared with the values directly calculated.

| Temperature (°K.) | Brightness $B$ (lightwatts cm.$^{-2}$) | $\log_{10} B$ (a) | $\log_{10} B$ calculated from eqn. (1) (b) | Difference (a)–(b) |
|---|---|---|---|---|
| 2000 | 0·0706 | $\bar{2}$·8490 | $\bar{2}$·8450 | +0·0040 |
| 2080 | 0·1137 | $\bar{1}$·0556 | $\bar{1}$·0538 | +0·0018 |
| 2160 | 0·1767 | $\bar{1}$·2473 | $\bar{1}$·2472 | +0·0001 |
| 2240 | 0·2665 | $\bar{1}$·4257 | $\bar{1}$·4268 | −0·0011 |
| 2320 | 0·3910 | $\bar{1}$·5922 | $\bar{1}$·5940 | −0·0018 |
| 2400 | 0·5595 | $\bar{1}$·7478 | $\bar{1}$·7500 | −0·0022 |
| 2480 | 0·7827 | $\bar{1}$·8936 | $\bar{1}$·8959 | −0·0023 |
| 2560 | 1·0730 | 0·0306 | 0·0328 | −0·0022 |
| 2640 | 1·443 | 0·1594 | 0·1613 | −0·0019 |
| 2720 | 1·910 | 0·2809 | 0·2823 | −0·0014 |
| 2800 | 2·486 | 0·3956 | 0·3964 | −0·0008 |
| 2880 | 3·192 | 0·5040 | 0·5042 | −0·0002 |
| 2960 | 4·040 | 0·6068 | 0·6061 | +0·0007 |
| 3040 | 5·050 | 0·7042 | 0·7026 | +0·0016 |
| 3120 | 6·262 | 0·7967 | 0·7942 | +0·0025 |

It is evident that the values of $B$ are given within ± 1% (an accuracy sufficient for many purposes) by the empirical equation (1).

The effect on the formula (1) of using other values for $C_1$ and $C_2$ may be noted. A change in $C_1$ produces a proportionate change in all the values of $J_\lambda d\lambda$, and hence in $\int_0^\infty J_\lambda V_\lambda\,d\lambda$, and so can be met by changing the constant 4·275 to

$[4{\cdot}275 + \log_{10}(C_1/3{\cdot}703)]$, where $C_1$ is expressed in erg sec.$^{-1}$ cm.$^{+2}\times 10^{-5}$. A change in $C_2$ affects the temperature-dependent term; since $\exp(C_2/\lambda T) \gg 1$ for the ranges of temperature and wave-length concerned, it is only necessary to change the constant 10860 to $(10860 \times C_2/14330)$, where $C_2$ is expressed in micron degrees. Thus if we adopt the recently suggested values of Birge (1941), namely $C_1 = 3{\cdot}7430 \times 10^{-5}$ erg sec.$^{-1}$ cm.$^{+2}$ and $C_2 = 14384{\cdot}8$ micron degrees, the relation becomes

$$\log_{10} B = 4{\cdot}279 - 10902/T.$$

## § 4. PHOTOMETRIC UNITS: THE RATIO OF THE NEW INTERNATIONAL CANDLE TO THE LIGHTWATT

It is possible to convert the above values of brightness directly from lightwatts per sq. cm. to "new international candles" per sq. cm., and to calculate the ratio of the new candle to the lightwatt, because the new candle is defined in terms of the intensity of a total radiator at a fixed temperature. This is of interest because the new candle is likely to come into force in the near future as the international unit of luminous intensity. By definition, the luminous intensity of a total radiator at the temperature of solidification of platinum is 60 new candles (International Lighting Vocabulary, C.I.E., 1938). The brightness in lightwatts per cm$^2$ of a total radiator at this temperature (taken here as 2045·9° K. (Wensel, 1937) is found, by interpolation from the directly calculated figures in the table above, to be 0·0932. The ratio of the new candle to the lightwatt is thus $60/0{\cdot}0932 = 643$ for the values of $C_1$ and $C_2$ used in computing the figures in the table; and to convert the values given above for the brightness of a total radiator to new candles per cm$^2$ it is only necessary to multiply each value by 643, or to add 2·808 to its logarithm.

## ACKNOWLEDGMENT

The author's thanks are due to Mr. J. S. Preston, of the National Physical Laboratory, for his advice and interest in this work.

*Note added in proof.* Mention should be made of the work of Ives (1926), who calculated values of the brightness of a total radiator at temperatures from 1200° K. to 10000° K. In these calculations, an approximate formula was used to express the C.I.E. relative-luminosity data, and errors of the order of 1% are likely. The author was unaware of this work when the above paper was written. He has also heen informed of a paper by W. Geiss (*Licht*, **12**, 33 (1943)), but has not been able to see a copy.

## REFERENCES

Birge, R. T., 1941. *Rev. Mod. Phys.* **13**, 233.
Hyde, E. P., Forsythe, W. E. and Cady, F. E., 1919. *Phys. Rev.* **13**, 45.
Ives, H. E., 1926. *J. Opt. Soc. Amer.* **12**, 75.
Judd, D. B., 1931. *Bur. Stand. J. Res., Wash.*, **6**, 465.
Nernst, W., 1906. *Phys. Z.* **7**, 380.
Skogland, J. F., 1929. *Misc. Pub. Bur. Stand., Wash.*, no. 86.
Wensel, H. T., 1937. *J. Res. Bur. Stand.* **6**, 1119.

# REVIEWS OF BOOKS

*Waveform Analysis*, by R. G. MANLEY. Pp. vii + 275 and 3 plates. (London: Chapman and Hall, Ltd., 1945.) 21*s.*

This book is not a treatise on harmonic analysis in the conventional sense but, as the author correctly emphasizes in the sub-title and preface, a practical guide to the rapid interpretation of wave records. Its primary aim is to explain, against a background of more formal analysis, methods, involving merely inspection and simple measurement, for estimating the frequency, amplitude and phase of the principal harmonic components of a large class of complex wave-forms which occur especially in vibration engineering. There are ten chapters, entitled: I, *Sine waves in combination*; II, *General properties of harmonic series*; III, *Basic analysis of recorded waveforms*; IV, *The envelope method*; V, *Method of superposition*; VI, *Fourier series: mathematical analysis*; VII, *Numerical methods*; VIII, *Mechanical and other aids to analysis*; IX, *Practical requirements for waveforms*; X, *Lissajous figures.*

Chapter I illustrates graphically and trigonometrically the synthesis of wave patterns (especially beats) from two or more sine components, and, together with Chapter II, provides that foundation of experience of the anatomy of wave-forms which is necessary for a ready diagnosis of harmonic content.

Chapter III seems a little dilute, and might be made more concise with advantage, but it serves to pave the way to the actual measuring of records and hence to Chapter IV, which forms the pivot of the book. This chapter, essentially diagnostic in character, shows how the construction of "envelopes", to touch the crests and troughs (or, more generally, to pass through any repeated characteristic kinks) of the undulations, makes it possible to identify the harmonics and to determine their characteristics. A lucid explanation is given of the surprising amount of information which may be extracted from a wave-form by an experienced interpreter without recourse to elaborate analysis. The method appears to have especial application to the routine examination of vibrograph, strain-gauge and similar records which are nowadays obtained in certain industries in numbers too great to permit of detailed analysis. A systematic procedure for applying the envelope method is developed in a number of tables somewhat reminiscent of chemical analysis.

For certain frequency ratios amongst the harmonics (such as 3 : 2 : 1), the envelope method fails. Chapter V shows how it is possible to meet these cases by dividing the complete cycle of the given wave into a number of equal parts which, by suitable summation and differencing, permit partial, and in favourable circumstances complete, separation of the harmonics.

It may be remarked that the envelope technique and similar simple methods, while of great service in resolving a wave composed of relatively few harmonics of comparable amplitude, do not promise so much help to the physicist, meteorologist and economist, who will still have to resort to Fourier, periodogram and correlogram analysis when confronted, as they frequently are, by waves having a long retinue of progressively diminishing harmonics.

The theoretical Chapter VI treats the formal representation of functions (including the usual examples of square and saw-tooth profiles) by Fourier series, and gives an adequate sketch of the mathematical background. While very properly leaving full rigour to the mathematicians, it does somewhat more than merely give plausible derivations of formulae. One cannot help regretting, however, that the nature of the book scarcely permits presentation of the connection between the author's "periodic existence functions" and their analogues, the Cauchy discontinuous functions. Gibbs' phenomenon at discontinuities is treated briefly, and Fourier integrals are mentioned but not applied.

Chapter VII gives a clear account of the numerical Fourier analysis of waves specified by equidistant ordinates, including computation schedules for schemes using 24 and 48 ordinates.

Chapter VIII describes in detail a versatile drawing-table type of mechanical harmonic analyser, and briefly mentions alternative devices, including electrical filter circuits. No attempt is made to survey the very rich field of mechanical and electrical aids which now exists.

General comments on the practical problems of recording and allowing for response distortion are given in Chapter IX, and the book closes with a chapter on the characteristics of two vibrations compared by means of Lissajous figures.

There are several appendices dealing with mathematical points, a list of elementary trigonometrical formulae, a useful table of sines and cosines adapted to facilitate harmonic analysis, a glossary of terms, a bibliography and an index.

A minor criticism of the presentation is that elementary details of mathematical manipulation are occasionally given more fully than is necessary if the reader is sufficiently mature to appreciate the chapter on Fourier analysis ; but this may encourage further reading, and it would therefore be beneficial if the bibliography were somewhat enlarged. Viewed as a whole, the book is well written and competently composed, and should be of considerable value to those concerned with vibration engineering. There are few misprints, of which the most obtrusive is the insertion of the terminal " s " of Lissajous in L'Hopital. The printing and binding are quite good. M. S. JONES.

*Introduction a l'étude des champs physiques*, by J. GRANIER. Pp. iv + 251. (Paris: Dunod, 1941.)

This book is not an introduction to *field theories* as now understood, but a general treatment of steady-state boundary problems where the differential equation is that of Laplace or of Poisson, that is of problems in electrostatics, magnetism, heat conduction, capillarity or hydrodynamics. The treatment is relatively elementary, and is confined in the main to problems in two dimensions, or those with spherical symmetry, so that there is no call to introduce Legendre functions, spherical harmonics or Bessel functions. The French nation has given the world some of its greatest geometers, and the tradition of geometrical reasoning has always been strong there, so that it is not surprising to note that, in this book, transformations from one co-ordinate system to another are seen from a rather more directly geometric aspect than they usually are in books written by Englishmen, who tend to see them as algebraic transformations.

After a preliminary chapter in which conjugate complex functions and systems of orthogonal curves are considered, the book is divided into chapters, each dealing with some phenomenon in which Laplacian fields present themselves, beginning with electrostatics. By taking the potential due to a point charge, and that due to an infinite straight line (a point charge in two dimensions), other cases are built up, the most elegant being that of a circular cylinder surrounding two line charges of opposite sign, a case which is found when considering the case of four line charges, by noting that there is one equipotential which is circular. A few cases are solved by means of conjugate functions, but these are generally introduced experimentally " to find out what cases they will represent ", rather than deliberately to solve particular problems posed beforehand. Even after the theorem of Schwartz on the behaviour of functions at branch points has been introduced, little use is made of it, though in fact it enables us to select with certainty the transformation appropriate to any plane problem with rectilinear boundaries. The chapter on magnetism finds its interest of course in doublets and other multiple charges, and does not carry the matter very far. The field due to a sphere in a field which would otherwise be uniform is worked out, but the case, so dear to most text-book writers, of an ellipsoid in a uniform field, is omitted. The misprints in the middle of page 121, where $R^3$ is omitted from the denominator of two successive equations, might mislead beginners. In the section on hydrodynamics we have such problems as the flow past a cylinder, and attention is then turned to cases where there is a free stream-line, as when a jet issues through a slit. No use is made here of the method devised by Kirchhoff of introducing a subsidiary plane in which the co-ordinates are the absolute magnitude and the direction of the velocity, but a great deal is extracted from the simple transformation, attributed to Joukowsky, $x=\frac{1}{2}(r+r^{-1})\cos\theta$ and $y=\frac{1}{2}(r-r^{-1})\sin\theta$, which analysts would think of as $w=(z+1/z)$. This, with its inverse transformation, and the addition of different fields of flow, yields an astonishing amount of information about flow round obstacles of practical interest, such as aerofoils.

The last quarter of the book is of a rather different nature, and considers the use of experimental methods for solving complicated problems—measurements of electrical resistance of diagrams cut out of metal foil, or tracing potential lines by means of probe measurements in an electrolyte, or plotting out a field by means of iron filings above a magnet, for example. The author has clearly had a very considerable experience with these methods, and is able to give advice on experimental arrangements which must be of immense value to any who propose to use the methods. Among them he includes the graphical method in which a network of orthogonal curves is drawn freehand, so spaced as to divide the field approximately into squares.

There is no doubt that the book will be of value to many, in showing the essential unity of the diverse branches of physics with which it deals, in enabling them to devise methods for attacking problems with which they are confronted, and as a book of reference, giving the solutions of many problems. When we consider that it was published in France in 1941, we can only imagine the difficulties which must have confronted author and publisher, and must admire them for producing the book at all ; that they produced such a good one must almost be a matter of wonder. J. H. A.

*The Quantum Theory of Radiation*, by W. Heitler. Second Edition. Pp. viii + 272. (London and Oxford : Sir Humphrey Milford, 1944.) 20*s*. net.

Ten years have passed since the first edition of this book appeared, and it has clearly established itself in the interval as an authoritative treatment of its extremely difficult subject. Although progress, particularly on the experimental side, has been steady, there has not yet been any spectacular advance which would call for a revision and re-writing of the theory, and in fact the present volume does not differ essentially from the first edition, save for interpolations here and there and an addition to the appendix.

One day, of course, a complete change of approach will be found, which will at one stroke give us a quantum theory completely in harmony with the postulates of relativity, and which will avoid the difficulties we find at present in connexion with the infinite self-energy of the point electron. Till this new formulation is made, it is difficult to imagine any more useful, or more instructive, text-book than this, though we may at times wish that someone would write a simple text to serve as an introduction to the subject. J. H. A.

*Tables of Elementary Functions*, by F. Emde. Pp. xii + 181. (Leipzig: Teubner, 1940; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) $3.20.

*Sechsstellige trigonometrische Tafeln*, by H. Brandenburg. Pp. xxiv + 304. (Leipzig : Lorentz, 1932 ; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) $5.00.

It is regretted that, in the reviews of these two books in the July issue of the *Proceedings* (this volume, pp. 368–370), no mention was made of the fact that the sole agent for these tables in Great Britain is *The Scientific Computing Service, Ltd.*, of 23 Bedford Square, London W.C. 1.